Encyclopedia of Geomorphology

Volume 1
A–I

Edited by A.S. Goudie

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

GEOMORPHOLOGY GEOMORPHOLOGIE

International Association of Geomorphologists

Illustrations

Figures

Plates

Tables

Foreword

As president of the International Association of Geomorphologists (IAG), a body that seeks to provide a forum for the promotion of geomorphology internationally, I am delighted that, in association with Routledge, it has been possible to produce this great encyclopedia. It is written by contributors from some thirty countries, all of whom have generously agreed that their royalties should go to the IAG. This will add substantially to the financial resources of the Association. The IAG is grateful to the editorial team, and in particular to Andrew Goudie, for the work they have done to bring it to fruition and I am sure that it will be an invaluable resource for the international geomorphological community.

Mario Panizza
Modena, Italy
October 2002

Preface

The term 'geomorphology' arose in the Geological Survey in the USA in the 1880s and was possibly coined by those two great pioneers, J.W. Powell and WJ McGee.

In 1891 McGee wrote: 'The phenomena of degradation form the subject of geomorphology, the novel branch of geology.' He plainly regarded geomorphology as being that part of geology which enabled the practitioner to reconstruct Earth history by looking at the evidence for past erosion, writing:

> A new period in the development of geologic science has dawned within a decade. In at least two American centres and one abroad it has come to be recognised that the later history of world growth may be read from the configuration of the hills as well as from the sediments and fossils of ancient oceans... The field of science is thereby broadened by the addition of a coordinate province – by the birth of a new geology which is destined to rank with the old. This is geomorphic geology, or geomorphology.

Of course, many scientists had studied the development of erosional landforms (see the magisterial history of Chorley *et al.* 1964) before the term was thus defined and since that time its meaning has become broader. Many geomorphologists believe that the purpose of geomorphology goes beyond reconstructing Earth history and that the core of the subject is the comprehension of the form of the ground surface and the processes which mould it. In recent years there has been a tendency for geomorphologists to become more deeply involved with understanding the processes of erosion, weathering, transport and deposition, with measuring the rates at which such processes operate, and with quantitative analysis of the forms of the ground surface (morphometry) and of the materials of which they are composed. Geomorphology now has many component branches and involves the study of a huge range of phenomena.

In 1968 Rhodes W. Fairbridge edited a large and invaluable encyclopedia of geomorphology that explored this diversity. However, geomorphology has changed greatly since that time, not least because of the plate tectonics paradigm, the revolution in our knowledge of the Quaternary Era brought about by new dating and environmental reconstruction techniques, the development of modelling and systems thinking, appreciation of the importance of organisms, application of geomorphology to the study of engineering problems and global change, a greater appreciation of the nature of geomorphological processes, and availability of a whole range of new technologies for analysis of data and materials, the development of satellite-borne remote sensing, and the exploration of space.

Over that time, due to the inspiration of the people to whom this book is dedicated, Geomorphology has for the first time organized itself internationally so that the geomorphological traditions that have grown up in different countries (see Walker and Grabau 1993) can interact as never before. It was therefore felt at the International Geomorphological Congress in Tokyo in August 2002 that the International Association of Geomorphologists (itself officially founded in 1989) should seek to publish a new and truly international Encyclopedia of Geomorphology that could survey the nature of the discipline at the turn of a new millennium. I am indebted to my Consultant Editors and the contributors from some thirty or so countries, who have made this endeavour possible.

Andrew S. Goudie
Oxford

References

Chorley, R.J., Dunn, A.J. and Beckinsale, R.P. (1964) *The History of the Study of Landforms*, Vol. 1, London: Methuen.

Fairbridge, R.W. (ed.) (1968) *Encyclopedia of Geomorphology*, New York: Reinhold.

McGee, W.J. (1891) The Pleistocene history of northeastern Iowa, *Eleventh Annual Report of the US Geological Survey*, 189–577.

Walker, H.J. and Grabau, W.E. (1993) *The Evolution of Geomorphology. A Nation-by-Nation Summary of Development*, Chichester: Wiley.

Thematic entry list

Bedform
Bedload
Bedrock channel
Beheaded valley
Blind valley
Bolson
Box valley
Braided river
Buried valley
Canyon
Cavitation
Channel, alluvial
Channelization
Comminution
Confluence, channel and river junction
Contributing area
Cross profile, valley
Cut-and-fill
Dambo
Debris torrent
Dell
Donga
Downstream fining
Drainage basin
Drainage density
Drainage pattern
Dry valley
Dune, fluvial
Estuary
Eversion
Fire
First order stream
Flash flood
Flood
Floodout
Floodplain
Flow regulation systems
Fluvial armour
Fluvial erosion quantification
Fluvial geomorphology
Glacideltaic
Glacifluvial
Gorge and ravine
Gravel-bed river
Ground water
Gully
Headward erosion
Hillslope-channel coupling
Hillslope hollow
Horton's Laws
Hydraulic geometry
Hydrological geomorphology
Hyperconcentrated flow
Initiation of motion
Inland delta
Interfluve
Knickpoint
Large woody debris
Levee
Long profile, river
Maximum flow efficiency
Meandering
Megafan
Mekgacha
Meltwater and meltwater channel
Mining impacts on rivers
Mobile bed
Mound spring
Outburst flood
Overflow channel
Overland flow
Oxbow
Palaeochannel
Palaeoflood
Palaeohydrology
Peat erosion
Pediment
Physical integrity of rivers
Piezometric
Pipe and piping
Point bar
Pool and riffle
Pot-hole
Prior stream
Quick flow
Raindrop impact, splash and wash
Rapids
Rejuvenation
Reynolds number
Rill
Riparian geomorphology
River capture
River continuum
River delta
River plume
River restoration
Runoff generation
Saltation
Sand-bed river
Scabland
Sediment load and yield
Sediment rating curve
Sediment routing
Sediment wave
Sedimentation
Sheet erosion, sheet flow, sheet wash
Sinuosity
Solute load and rating curve
Step-pool system
Stream ordering
Stream power
Stream restoration
Subcutaneous flow
Suffosion
Suspended load
Terrace, river
Underfit stream
Valley
Valley meander
Wadi
Waterfall
Watershed
Yazoo

Glacial

Arête
Bergschrund
Calving glacier
Cirque, glacial
Deglaciation
Diamictite
Drumlin
Equilibrium line of glaciers
Erratic
Esker
Fjord
Glaciaeolian
Glacial deposition
Glacial erosion
Glacial isostasy
Glacial protectionism
Glacial theory
Glacideltaic
Glacier
Glacifluvial
Glacilacustrine
Glacimarine
Glacipressure
Glacitectonic
Glacitectonic cavity
Hanging valley
Ice
Ice ages
Ice sheet
Ice stagnation topography

Ice stream
Iceberg
Ice dam, glacier dam
Kame
Kettle and kettle hole
Mass balance of glaciers
Meltwater and meltwater channel
Moraine
Moulin
Neoglaciation
Nunatak
Overflow channel
Paraglacial
Pinning point
Pot-hole
Pressure melting point
Proglacial landform
Regelation
Roche moutonnée
Rock glacier
Sastrugi
Sichelwanne
Striation
Subglacial geomorphology
Supraglacial
Surging glacier
Trimline, glacial
Tunnel valley
Urstromtäler

Hazards and environmental geomorphology

Applied geomorphology
Arroyo
Avalanche, snow
Beach nourishment
Catastrophism
Channelization
Dam
Debris flow
Debris torrent
Desertification
Dust storm
El Niño effects
Engineering geomorphology
Environmental geomorphology
Expansive soil
Factor of safety
Failure
Flash flood
Flood
Flow regulation systems
Geoindicator
Geomorphological hazard
Geosite
Global warming
Groyne
Hydrocompaction
Ice dam, glacier dam
Integrated coastal management
Lahar
Landslide
Landslide dam
Liquefaction
Managed retreat
Mass movement
Mining impacts on rivers
Nuée ardente
Outburst flood
Quickclay
Quicksand
River restoration
Rockfall
Rocky desertification
Soil conservation
Soil erosion
Stream restoration
Sturzstrom
Subsidence
Surging glacier
Tsunami
Urban geomorphology

Karst

Biokarst
Blue hole
Cave
Cavernous weathering
Cenote
Corrosion
Cryptokarst
Dissolution
Doline
Dye tracing
Endokarst
Epikarst
Gypsum karst
Karren
Karst
Limestone pavement
Micro-erosion meter
Palaeokarst and relict karst
Pan
Polje
Pseudokarst
Rocky desertification
Salt karst
Speleothem
Spring, springhead
Subsidence
Syngenetic karst
Tufa and travertine
Turlough
Volcanic karst

Lacustrine

Alas
Cenote
Dam
Daya
Glacilacustrine
Ice dam, glacier dam
Lagoon, coastal
Lake
Landslide dam
Oriented lake
Oxbow
Pan
Paternoster lake
Pluvial lake

Palaeogeomorphology

Base level
Buried valley
Chronosequence
Climato-genetic geomorphology
Cosmogenic dating
Cycle of erosion
Dating methods
Denudation chronology
Dendrochronology
Dendrogeomorphology
Divergent erosion
Etching, etchplain and etchplanation
Eustasy
Exhumed landform
Fission track analysis
Geomorphic evolution
Glacial theory
Grade, concept of
High-energy window
Holocene geomorphology
Ice Ages

Periglacial

Slopes and mass movements

Soils and materials

Structural

Techniques

Tectonic and volcanic

Weathering

A

ABRASION

The mechanical wearing down, scraping, or grinding away of a rock surface by friction, ensuing from collision between particles during their transport in wind, ice, running water, waves or gravity. The effectiveness of abrasion depends upon the concentration, hardness and kinetic energy of the impacting particles, alongside the resistance of the bedrock surface. Abrasion may scour, polish, scratch or smooth existing rock faces. Abrasion ramps are seaward sloping platforms (typically 1° gradient) formed at the base of cliffs in intertidal environments due to continued wave abrasion.

Further reading

Hamblin, W.K. and Christiansen, E.H. (2001) *Earth's Dynamic Systems*, 9th edition, Upper Saddle River, NJ: Prentice Hall.

STEVE WARD

ACCRETION

The gradual enlargement of an area of land through the natural accumulation of sediment, washed up from a river, lake or sea. Sediment accretion is the basic process of wetland formation, as continuous flooding and subsequent receding river flows emplace sediment which then provides the soil base for wetlands.

Accretion also refers to the theory that continents have increased their surface area during geological history by the addition of marine sediments at their boundaries via tectonic collision with other oceanic or continental plates.

Further reading

Pye, K. (1994) *Sediment Transport and Depositional Processes*, Oxford: Blackwell Scientific.

STEVE WARD

ACTIVE AND CAPABLE FAULT

Currently no universally accepted definition has been agreed upon for 'active fault', nor have the principles and criteria for the identification of active faults and their ranking been worked out. As a result, the various definitions of fault activity terms are the source of some confusion and discussion, both in literature and in practice (see FAULT AND FAULT SCARP).

An important review on this topic was presented by Slemmons and McKinney (1977) who, after examining numerous papers, suggested the following definitions. An 'active fault' is a fault that has slipped during the present seismotectonic regime and is therefore likely to show renewed displacement in the future. Fault activity may be indicated by historical, geological, seismological, geodetic or other geophysical evidence. The definitions for 'capable faults', which were specified for siting nuclear reactors, restrict this term to faults that have been displaced once during the past 35,000 years, or movements of a recurring nature within the past 500,000 years or faults which have been active during the Late Quaternary.

The term active fault in Japan was defined as a fault which has moved repeatedly in recent geological times and could resume activity in the future. Subsequently, this term has been used for faults which have moved during the Quaternary. The analysis of topographic features has provided the most important clues in the work of recognizing active faults (RGAFJ 1980).

A particular definition was given by Panizza and Castaldini (1987) who distinguish two categories: (1) active fault: proven displacement of rocks and/or significant forms; (2) fault held to be active: on the basis of supporting geomorphological or other evidence, but showing no visible displacement of rock or other significant forms. Rocks and/or 'significant' landforms are those included in the neotectonic period considered. The distinction between 'active' and 'held to be active' faults is finalized to constrain in a more precise and less subjective way the concept of fault activity.

The 'World Map of Major Active Faults' shows five fault age categories (historical to <1.6 Ma). Slip rate, which is used as a proxy for fault activity, is classified in four categories ranging from $<0.2\,\mathrm{mm\,year^{-1}}$ to $>5\,\mathrm{mm\,year^{-1}}$. The maps are accompanied by a database which describes evidence for Quaternary faulting, geomorphic expression and paleoseismic parameters (Trifonov and Machette 1993).

In some glossaries, an active fault is defined as 'A fault along which there is recurrent movement, which is usually indicated by small, periodic displacements or seismic activity' (Bates and Jackson 1987), or as 'A fault likely to move at the present day' (Ollier 1988).

A paper on the most commonly used terms associated with seismogenetic faults in the United States was published by Machette (2000). The author notes that the three following terms are used in a variety of ways and for different reasons or applications:

- Active fault: one demonstrating current movement or action (what is meant by 'current'? Contemporary, historical, Holocene or Quaternary?).
- Capable fault: one having the capability for movements.
- A potentially active fault: one capable of being or becoming active (this definition is very similar to that of capable fault).

On the Internet various definitions pinpointing the indeterminateness of the term can be found, such as:

1 The definition of active fault is not straightforward. In some cases, the maximum age that can be determined by means of Carbon 14 analyses (35,000–50,000 years) is used as a time span for such measurements: if a fault can be shown not to have been active within this time span then it is not active (http://www.geol.binghamton.edu/class/geo205/html/faults.html).
2 Active faults are structures along which displacements are expected to occur. By definition, since a shallow earthquake is a process that produces displacement across a fault, all shallow earthquakes occur on active faults (http:// www.eas.slu.edu/People/CJAmmon/HTML/Classes/introQuakes/Notes/faults.html).
3 'A fault that is likely to undergo displacement by another earthquake sometime in the future.' Faults are commonly considered to be active if they have moved one or more times in the last 10,000 years (http://earthquake.usgs.gov/image_glossary).
4 An active fault is one that has moved at least once in geologically recent times. In the Californian definition it means a movement occurring within the last 11,000 years, rather than the longer period of 125,000 years used on New Zealand maps (http://www.gsnz.org.nz/gsprfa.htm).

In short, on the concepts of 'active fault' and 'capable fault' the following remarks can be made:

1 the terms are used to indicate faults which have been subject to movement in recent geological time or which might move at present or in the future;
2 their age limits vary depending on the authors;
3 active faults are often associated with strong earthquakes.

Identification of active and capable faults can be based on direct and/or indirect criteria: historical, geological, geomorphological, geomorphic, seismological, geodetic, geochemical, geophysical and volcanic.

Finally, apart from the terminological aspects, some of the major active faults around the world include: the North Anatolian fault in Turkey, the Dead Sea Valley between Israel and Jordan, the Philippine fault, the San Andreas fault in California, the Red River fault in China and the South Island alpine fault in New Zealand.

References

Bates, R.L. and Jackson, J.A. (1987) *Glossary of Geology*, American Geological Institute, Alexandria, VA.
Machette, M.N. (2000) Active, capable and potentially active faults: a paleoseismologic prospective, *Journal of Geodynamics* 29, 387–392.
Ollier, C.D. (1988) *Glossary of Morphotectonics*, 3rd edition, Dept of Geography and Planning, University of New England, Armidale, Australia.
Panizza, M. and Castaldini, D. (1987) Neotectonic research in applied geomorphological studies, *Zeitschrift für Geomorphologie Supplementband* 63, 173–211.
RGAFJ (The Research Group for Active Faults of Japan) (1980) Active faults in and around Japan: distribution and degree of activity, *Journal of Natural Disaster Science* 2(2), 61–99.
Slemmons, D.B. and McKinney, R. (1977) Definition of 'active fault', *US Army Engineer Waterways Experiment Station, Soils and Pavements Laboratory, miscellaneous paper S*, 77–8, Vicksburg.
Trifonov, V.G. and Machette, M.N. (1993) The world map of major active faults, *Annali di Geofisica* 36(3–4), 225–236.

DORIANO CASTALDINI AND DORINA CAMELIA ILIES

ACTIVE LAYER

Ground above PERMAFROST which thaws in summer and freezes again in winter. In the northern hemisphere, it reaches its full depth each year in late August or September. The active layer is critical to the ecology of permafrost terrain, as it provides a rooting zone for plants and is a seasonal aquifer. An ice-rich zone, commonly below the base of the active layer, is responsible for the sensitivity of permafrost terrain to disturbance. Deepening of the active layer and melting of the ground ice leads to subsidence in flat terrain, and landslides with accelerated erosion on slopes (Mackay 1970).

The active layer thaws once air temperature is above 0 °C and the snow cover has melted. The total depth depends on the length and surface temperature of the thawing season, the ice content of the ground, the thermal conductivity of soil materials, and the temperature of near-surface permafrost. The active layer is thickest in bedrock, where there is little ice to melt. In unconsolidated sediments, the thickness is greatest in dry, sandy soils or gravel, where the depth may be enhanced by heat advected in groundwater, and thinnest in peat. Local variation in soil materials may be reflected in active-layer depth, as in hummocky terrain, where the base of the active layer forms a mirror image of the ground surface, with depth greatest beneath the mineral-soil centres and least beneath the organic-rich circumference of the hummocks.

At the end of the thaw season, freezing of the active layer usually begins from the bottom upwards. Upfreezing commonly accounts for up to 10 per cent of the thickness. During upfreezing, moisture is drawn downward into permafrost from the base of the active layer, leading to development of the ice-rich zone. Simultaneously, soil water is drawn upwards from the rest of the active layer, to freeze near the ground surface. As a result, the centre of the active layer tends to be dry when frozen. Stones and structures embedded in the active layer may be pulled upwards as the ground freezes. Characteristically these objects are supported from below during thawing the following summer, leading to their progressive jacking out of the ground.

Freezing and thawing of the active layer modifies the annual propagation of surface temperature into permafrost. Cooling of permafrost in autumn is delayed by freezing, which may take several months, depending primarily on the water content and snow cover. Mean annual temperature decreases with depth in the active layer, due to the seasonal difference in soil thermal properties wrought by freezing and thawing. The difference in mean annual temperature, or thermal offset, between the ground surface and the top of permafrost may be over 2 °C, increasing with water content and depth of active layer (Romanovsky and Osterkamp 1995).

References

Mackay, J.R. (1970) Disturbances to the tundra and forest tundra environment of the western Arctic, *Canadian Geotechnical Journal* 7, 420–432.
Romanovsky, V.E. and Osterkamp, T.E. (1995) Interannual variations of the thermal regime of the active layer and near-surface permafrost in northern Alaska, *Permafrost and Periglacial Processes* 6, 313–335.

C.R. BURN

ACTIVE MARGIN

In plate tectonic theory ocean crust is created by SEAFLOOR SPREADING, and old crust is consumed at subduction sites. A continental margin where subduction occurs is called an active continental margin. Active margins occupy essentially the borders of the Pacific: the west Pacific borders are

ISLAND ARC type; the western margins of the Americas are the other type, to be described here.

The spreading of the Atlantic causes America to move west, where it overrides the seafloor, which is subducted. Plate collision is thought to fold and uplift the continental edge to form mountains and their internal structures, and also create a deep trench offshore where sediments are deposited. These may be scraped off the subducted slab to form an accretionary prism, or subducted where they may produce granites, and andesitic magma which erupts as volcanoes.

The Pacific border of the Americas falls into three main units with different MORPHOTECTONICS: South America, Central America and North America.

The Andes run along the entire western side of South America, divided for most of their length into Eastern and Western Cordilleras, with a graben between called the Inter-Andean Depression. Bedrock is folded and faulted Palaeozoic and Mesozoic rocks, with granite intrusions. The region was largely eroded to a plain before ignimbrites spread over large areas, and planation was complete in the Neogene. The area was uplifted as linear fault blocks in the Plio-Pleistocene, or earlier in some places. The large strato-volcanoes are of Quaternary age, erupted onto the planation surface. Major thrust faults diverge from the centre of the Andes in a symmetrical way, hard to explain by one-sided subduction.

Offshore a deep trench extends as far north as Mexico. Trenches have many graben and normal faults indicating extension. Sediments are usually horizontal, and some trenches are almost empty. Mesozoic plutons constitute the world's greatest granite batholith which runs the length of the Andes, covering 15 per cent of the Andes surface. The alignment parallel to the coast suggests some control on the location and possibly the origin of the Andes, but the plutons took over 70 million years to rise and intrusion ceased about 30 million years ago, long before the uplift of the Andes (Gansser 1973).

The many great volcanoes found along the Andes (with some gaps, and some double lines) are Quaternary. They are on the top of horsts, usually close to the Inter-Andean Depression.

Central America can be regarded as the Middle America arc. The trench has no accretionary prism, and sediments are horizontal.

A basement of metamorphic rocks and granites is exposed in northern Honduras. This is block faulted, and split by the Honduras Depression consisting of north–south graben that opened in the early Pliocene. The chain of volcanoes close to the south coast consists of five straight-line segments. These young cones are built on a basement of older volcanics. The same basement forms the Nicaraguan volcanic upland, separated from the young volcanoes by a major fault scarp. To the north these volcanics overlap the Honduras Massif.

The Isthmian link to Panama is not the young volcanic chain or even rocks of the Nicaraguan volcanic upland, but consists of even older volcanics. Block faulting is common.

Western North America is largely a collage of exotic terranes (Howell 1989). There are abundant strike-slip faults (such as the San Andreas fault) with movement of hundreds of kilometres. There is no offshore trench, but an offshore topography of basins and swells, possibly related to strike-slip fault blocks. These differences perhaps occur because the mid-ocean ridge runs aground near the Mexico/USA border. To the north the transform faults associated with seafloor spreading affect the continental margin as they run nearly parallel to it.

The Pacific border region of the USA consists of two main ranges: in the west are the Coast Ranges, in the east to the north are the Cascades and to the south the Sierra Nevada. The Coast Range seems to have formed as a large but rather simple arch. The Cascade Range is mainly a huge pile of volcanic rocks, with many famous strato-volcanoes such as Mount Shasta and Mount St Helens. The Sierra Nevada is a huge tilt block, mainly uplifted in the Quaternary. The Coast Ranges of Canada are a continuation of the Cascades of the United States, and also consist of a simple arch.

Planation surfaces are common on the North American cordillera (Ollier and Pain 2000). The mountains of North America were uplifted in the Neogene, mostly within the past 5 million years, though subduction has presumably been going on for the life of the Pacific, at least 200 Ma.

Plate tectonic theory has been applied not only to coastal ranges, but to mountains 1,500 km inland (Miller and Gans 1997). The Rocky Mountains consist of elongated blocks aligned in all directions, including east–west (Uinta Mountains).The blocks have Precambrian cores, and divergent thrust faults on both sides. They are too far inland to be explained by subduction, separated from the Pacific by the extensional Basin and Range Province, and the uplift occurred in the last few million years.

As Gansser (1973) explained, plate tectonic theories that use the Andes as a model adopt simplified assumptions which neglect the fact that only the recent morphogenic uplift made the apparently uniform Andes, masking a very complicated geological history. The same seems true of North and Central America.

References

Gansser, A. (1973) Facts and theories on the Andes, *Journal of the Geological Society of London* 129, 93–131.

Howell, D.G. (1989) *Tectonics of Suspect Terranes: Mountain Building and Continental Growth*, London: Chapman and Hall.

Miller, E.L. and Gans, P.B. (1997) The North American Cordillera, in B.A. van der Pluijm and S. Marshak (eds) *Earth Structure: An Introduction to Structural Geology and Tectonics*, 424–429, New York: WCB/McGraw-Hill.

Ollier, C.D. and Pain, C.F. (2000) *The Origin of Mountains*, London: Routledge.

Further reading

Leggett, J.K. (1982) *Trench-Forearc Geology: Sedimentation and Tectonics in Modern and Ancient Active Plate Margins*, Oxford: Blackwell.

Nairn, A.E.M., Stehli, F.G. and Uyeda, S. (1985) *The Ocean Basins and Margins. Vol. 7A The Pacific Ocean*, New York: Plenum.

SEE ALSO: mountain geomorphology; plate tectonics

CLIFF OLLIER

ACTUALISM

Actualism is a concept based on the premise that present causes of environmental change are sufficient to explain events of the past. Causes of changes in the past differ not in kind, but often in energy, from those now in operation. The French term *actualisme* and the German terms *aktualismus* or *aktualitatsprinsip* are commonly used in Europe in opposition to catastrophism. Hooykaas (1970) makes a distinction between actualistic methodology and actualistic historical description. Tidal variation over geological time provides an instructive example. Actualist methodology leads to the conclusion that the Moon and Earth were very much closer and that gravitational attraction was therefore greater before 3.5 billion years BP. Huge tidal ranges require a catastrophist historical description.

Reference

Hooykaas, R. (1970) *Catastrophism in Geology: Its Scientific Character in Relation to Actualism and Uniformitarianism*, Amsterdam: North-Holland Publishing Co.

SEE ALSO: catastrophism; uniformitarianism

OLAV SLAYMAKER

ADHESION

Adhesion refers to the adhering of wind-blown sand to a wet or damp surface. Adhesion is most common in damp or wet INTERDUNES between active dunes, but also occurs on SANDSHEETS, beaches, riverbanks and damp portions of dunes. Adhesion ripples and plane bed are the most common surface features that result from adhesion, and each forms a distinctive sedimentary structure with deposition. A related feature is formed by adhesion of sediment to salt during periods of high humidity.

Further reading

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

GARY KOCUREK

AEOLATION

The moulding of desert landscapes by the erosional action of wind (see WIND EROSION OF SOIL). At the start of the twentieth century there was a phase of what has been termed 'extravagant aeolation' (Cooke and Warren 1973). This had its roots in the work undertaken in Africa by French and German geomorphologists, such as Walther and Passarge, but was put forward in its most exuberant form in the USA by Keyes (1912), who believed that material weakened by thermoclasty (INSOLATION WEATHERING) would be evacuated by wind and deposited as dust sheets on desert margins. He argued that the end result of such activity would be the formation of great plains, mountain ranges without foothills, and towering eminences. As he remarked (p. 551):

> Under conditions of aridity plain meets mountain sharply. The bevelled rock-floor of many intermont plains throughout the dry regions is explicable on no known activity of water action in such situations. Existence of isolated plateau

plains rising abruptly out of the general plains surface far from any sight of running water is an anomaly met with only in the desert.

Not all American geomorphologists took such a firm view as Keyes, and Tolman (1909), for instance, in his study of the Arizona BOLSON region, considered the role of both fluvial and aeolian processes to be important and recognized that STONE PAVEMENTS 'fortified' large tracts of the arid region of the south-west of the USA against wind attack.

Aeolianist views declined in popularity so that from about 1920 onwards the belief that entire landscapes were shaped by wind became less acceptable. The reasons for the decline of aeolianist views were many.

First, the great PEDIMENT landscapes of the American deserts were seen, following the work of McGee (1897) and others, as being attributable to planation by sheetflood activity. The second reason for the decline of aeolianist views was that many desert landscapes were thought to have been moulded by fluvial processes that had been more powerful and widespread during the pluvial phases that were held to be a feature of the Pleistocene. Third, doubt was expressed about the power of thermoclasty as a process capable of preparing desert surfaces for subsequent aeolian attack. Such doubt largely arose because of laboratory simulations. Fourth, it was widely held that lag gravels (stone pavements) and salt and clay crusts would limit the extent to which aeolian processes could cause excavation of surfaces below the water table. Fifth, it became apparent that many of the world's great LOESS deposits, in North America, China and the erstwhile USSR, were the product of deflation from glacial areas rather than from deserts. Glacial grinding was thought to be the most efficient way of producing silt-sized quartz particles. Sixth, it was recognized that not all deserts had either adequate supplies of abrasive sand or of frequent high-velocity winds for wind erosion to be achieved with any degree of facility. Finally, features that were conceded to have an aeolian origin (e.g. YARDANGS, VENTIFACTS and pedestal rocks) were thought to be but minor, bizarre embellishments of otherwise fluvial environments, whilst other possibly aeolian features (notably stone pavements and closed depressions) were also explicable by other means. STONE PAVEMENTS, for example, could be the product of the removal of fine sediments by sheetflood activity or they could result from vertical sorting processes associated with wetting and drying, dust inputs, salt hydration or freezing and thawing. Deflational removal of fines to leave a lag was just one possible formation mechanism. In the same way, closed depressions could be attributed to wind excavation, but might also be explained by tectonic, solutional or zoogenic processes.

Nevertheless, the power of wind erosion cannot be dismissed. Closed depressions (PANs) and wind moulded landforms (yardangs) are important landforms in some arid areas and wind erosion plays a significant role in their development.

References

Cooke, R.U. and Warren, A. (1973) *Geomorphology in Deserts*, London: Batsford.

Keyes, C.R. (1912) Deflative scheme of the geographic cycle in an arid climate, *Geological Society of America Bulletin* 23, 537–562.

McGee, W.J. (1897) Sheetflood erosion, *Geological Society of America Bulletin* 8, 87–112.

Tolman, C.F. (1909) Erosion and deposition in the southern Arizona bolson region, *Journal of Geology* 17, 126–163.

A.S. GOUDIE

AEOLIAN GEOMORPHOLOGY

Aeolian geomorphology is the study of the effect of the wind on Earth surface processes and landforms. It encompasses studies of the fundamental physical mechanisms and movement of materials at the scale of a single grain, studies of the development of landforms such as dunes (see DUNE, AEOLIAN) and YARDANGs, and studies of the wider effect of the wind at the regional scale of SAND SEA AND DUNEFIELD, SANDSHEETs and LOESS deposition. It is also concerned with applied aspects of aeolian activity. In recent years this has led to a particular focus on the erosion by wind of soils in agricultural lands, especially in the semi-arid lands (see WIND EROSION OF SOIL and DEFLATION). Aeolian geomorphology also includes the study of the palaeoenvironmental significance of aeolian features, for landscapes can be just as sensitive to changes in the activity of the wind as they are to water and ice. As with other elements of geomorphology, technological changes in recent years have led to considerable advances in the understanding of aeolian geomorphology.

At the scale of the movement of individual sand and dust particles, the benchmark work was undertaken by Bagnold and summarized in his *Physics of Blown Sand and Desert Dunes* (1941). Since Bagnold's work very considerable progress has been made in the use of wind tunnels and field instruments (see WIND TUNNELS IN GEOMORPHOLOGY). Although the fundamental physics of aeolian sand transport is much as Bagnold described it, very considerable detail has been added in recent years (see also AEOLIAN PROCESSES and SALTATION).

At the scale of individual landforms considerable progress has also been made. As a result of the improvement of technologies for data capture there has been a spate of studies of wind flow and sand flux on single dunes (e.g. Tsoar 1983; Walker 1999) reviewed by Wiggs (2001). Often these are coupled with improving surveying techniques that enable accurate measurement of change (e.g. Stokes *et al.* 1999). In addition, ground penetrating radar (GPR) is now being routinely used to ascertain the internal sedimentary structure of dunes as an important indication of the evolutionary history of dunes (e.g. Bristow *et al.* 2000). Studies have also investigated erosional features such as YARDANGs and VENTIFACTs (e.g. Laity 1994).

At the regional scale the development of remote sensing has enabled a better grasp of the relationships between landforms. The advance of remote sensing investigations in dryland areas has been of particular importance because desert areas are often difficult to access. The pioneering work of McKee and co-workers (McKee 1979) has been followed by numerous applications of remote sensing in aeolian studies. Imagery has been used to map dune patterns (e.g. Al-Dabi *et al.* 1997), detect small changes in dune morphology using high resolution synthetic aperture radar (SAR) imagery (e.g. Blumberg 1998), map and detect dust emissions from dryland pan systems (e.g. Eckardt *et al.* 2001) and detect mineral assemblages (e.g. White *et al.* 1997).

Aeolian features also hold considerable palaeoenvironmental information because aeolian activity is sensitive to changes in environmental controls such as wind energy and moisture availability. The extent of dunefields at the last glacial maximum was used as a surrogate indicator of global aridity by Sarnthein (1978) but a basic on/off classification of aeolian activity is now seen as too simplistic (Livingstone and Thomas 1993). Kocurek and Lancaster (1999), for instance, have sought to incorporate variability of sediment availability along with wind energy in discussions of past aeolian activity in the Mojave Desert.

A profound impact on aeolian studies has been the development of luminescence dating techniques (see DATING METHODS). Many aeolian deposits lack organic matter and so have not been susceptible to radiocarbon dating. Since the early 1980s luminescence dating primarily of quartz grains has enabled dating of aeolian deposits such as dunes and loess, and luminescence dates in aeolian studies are now commonplace (e.g. Stokes *et al.* 1997).

Improvement of dating techniques has led to considerable interest in the palaeoenvironmental information stored in LOESS (terrestrial deposits of aeolian dust). The best documented of these are the deposits of the Chinese loess plateau. Here mineral magnetism has been used as a proxy for weathering of PALAEOSOLs, patterns of magnetic reversals have been used to date deposits covering the past 2.5 million years and loess particle size has been used as an indicator of palaeo wind speeds. Techniques developed on the Chinese deposits have been extended to loess deposits elsewhere and knowledge of the extent and nature of world loess deposits has steadily increased (e.g. Derbyshire 2001).

Aeolian geomorphology has moved on considerably since the claims of Keyes in the early part of the twentieth century (see AEOLATION). The task that faces aeolian geomorphologists is to move from studies of individual landforms formed predominantly by aeolian activity to consider the wider role of the wind alongside water and ice in forming landscapes (e.g. Bullard and Livingstone 2002).

References

Al-Dabi, H., Koch, M., Al-Sarawi, M. and El-Baz, F. (1997) Evolution of sand dune patterns in space and time in north-western Kuwait using Landsat images, *Journal of Arid Environments* 36, 15–24.

Bagnold, R.A. (1941) *The Physics of Blown Sand and Desert Dunes*, London: Methuen.

Blumberg, D.G. (1998) Remote sensing of desert dune forms by polarimetric synthetic aperture radar (SAR), *Remote Sensing of the Environment* 65, 204–216.

Bristow, C.S., Bailey, S.D. and Lancaster, N. (2000) The sedimentary structure of linear sand dunes, *Nature* 406, 56–59.

Bullard, J.E. and Livingstone, I. (2002) Interactions between aeolian and fluvial systems in dryland environments, *Area* 34, 8–16.

Derbyshire, E. (2001) Recent research on loess and palaeosols, pure and applied: a preface, *Earth-Science Reviews* 54, 1–4.

Eckardt, F., Drake, N., Goudie, A.S., White, K. and Viles, H. (2001) The role of playas in pedogenic gypsum crust formation in the central Namib desert: a theoretical model, *Earth Surface Processes and Landforms* 26, 1,177–1,193.

Kocurek, G. and Lancaster, N. (1999) Aeolian system sediment states: theory and Mojave Desert Kelso Dune field example, *Sedimentology* 46, 505–515.

Laity, J.E. (1994) Landforms of aeolian erosion, in A.D. Abrahams and A.J. Parsons (eds) *Geomorphology of Desert Environments*, 506–535, London: Chapman and Hall.

Livingstone, I. and Thomas, D.S.G. (1993) Modes of linear dune activity and their palaeoenvironmental significance: an evaluation with reference to southern African examples, in K. Pye (ed.) *The Dynamics and Environmental Context of Aeolian Sedimentary Systems*, Geological Society Special Publication 72, 91–101, London: Geological Society.

McKee, E.D. (ed.) (1979) A study of global sand seas, *United States Geological Survey Professional Paper* 1052.

Sarnthein, M. (1978) Sand deserts during glacial maximum and climatic optimum, *Nature* 272, 43–46.

Stokes, S., Thomas, D.S.G. and Washington, R. (1997) Multiple episodes of aridity in southern Africa since the last interglacial period, *Nature* 388, 154–158.

Stokes, S., Goudie, A.S., Ballard, J., Gifford, C., Samieh, S., Embabi, N. and El-Rashidi, O.A. (1999) Accurate dune displacement and morphometric data using kinematic GPS, *Zeitschrift für Geomorphologie Supplementband* 116, 195–214.

Tsoar, H. (1983) Dynamic processes acting on a longitudinal (seif) dune, *Sedimentology* 30, 567–578.

Walker, I.J. (1999) Secondary airflow and sediment transport in the lee of a reversing dune, *Earth Surface Processes and Landforms* 24, 437–448.

White, K., Walden, J., Drake, N., Eckardt, F. and Settle, J. (1997) Mapping the iron oxide content of dune sands, Namib Sand Sea, Namibia, using Landsat Thematic Mapper data, *Remote Sensing of the Environment* 62, 30–39.

Wiggs, G.F.S. (2001) Desert dune processes and dynamics, *Progress in Physical Geography* 25, 53–79.

Further reading

Goudie, A.S., Livingstone, I. and Stokes, S. (eds) (1999) *Aeolian Environments, Sediments and Landforms*, Chichester: Wiley.

Livingstone, I. and Warren, A. (1996) *Aeolian Geomorphology*, London: Longman.

IAN LIVINGSTONE AND GILES F.S. WIGGS

AEOLIAN PROCESSES

Wind is the movement of the mixture of gases that constitute the air. It is a fluid like water and obeys the same fundamental physical mechanisms as water. However, there are clear distinctions between the effects of water and wind at the Earth's surface. Air is 100 times less dense than water and consequently is only able to carry small clastic material. However, wind is not constrained by channels in the way that much water action is and consequently its influence can be much wider spread. Paradoxically, it is this wide spread of activity that means that aeolian activity sometimes goes unrecognized. A few millimetres of erosion or deposition of material over a large area is much less obvious than the erosion of rills and gullies or deposition of bars in channels even though the total amount of material moved may be similar.

Controls on aeolian processes

While aeolian activity is often associated with hot deserts, it is not restricted to these areas, although these are among the regions with the most favourable conditions for aeolian activity. Primary requirements for aeolian activity are: sufficient wind energy; material of a size that can be transported; and surface conditions that make that material available to the wind. Aeolian activity is therefore controlled by transport capacity, sediment supply and sediment availability (Kocurek and Lancaster 1999).

- *Transport capacity* Most places on the Earth's surface experience sufficient wind energy for aeolian processes to operate, so wind energy is rarely a limiting factor. The high levels of aeolian activity in low-latitude deserts do not occur because they are windier than other places. In fact the windiest places on Earth are close to the poles and around coastlines.
- *Sediment supply* Because wind is not as dense or viscous as water it is much more selective about the size of material that it can carry. The size of material most readily entrained by the wind is fine sand (see below). Wind rarely carries material above sand size, although transport of gravel-sized particles (>2 mm) has been reported from the dry valleys in Antarctica where wind speeds are very high and the extremely cold air is dense. The size-selectivity of wind means that surface materials must usually be sand- or dust-sized to be entrained. Often this requires that the materials are pre-sorted by other fluvial, glacial or marine processes. Some of the best sources of aeolian material are alluvial fans, glacial outwash plains and beaches (Bullard and Livingstone 2002).

- *Sediment availability* Provided with sufficient wind energy and material of the right size, the remaining control on aeolian processes is the surface conditions. Deserts have high levels of aeolian activity because soil, vegetation or moisture do not seal the surface. Conversely, aeolian activity is more rare in mid-latitudes, not because of lack of wind or material of the right size, but because surface conditions prevent the wind from entraining material.

Processes of wind erosion

Erosion of materials at the Earth's surface by the wind occurs as a result of two processes: deflation and abrasion (both of these mechanisms also occur in flowing water although the equivalent term 'fluid stressing' is used instead of deflation in fluvial geomorphology). DEFLATION is very simply the entrainment of material by the wind. Surfaces on which dust- or sand-sized material are exposed are particularly susceptible and in some places agricultural land with sandy or dusty soils where farmers expose the soil by ploughing is subject to considerable deflation by the wind (see WIND EROSION OF SOIL). Abrasion is caused by bombardment by particles being transported by the wind, most usually by SALTATION (see below). The impact of these transported grains can cause considerable sculpting of natural and built features. YARDANGS and VENTIFACTS are the geomorphological features most affected by abrasion.

Sediment entrainment

Aeolian sediment entrainment on a stable non-eroding surface occurs when the shear stress of the wind (a function of wind speed, turbulent energy and surface roughness) overcomes forces of particle cohesion, packing and weight. The principal erosive forces include lift, form drag and surface drag. The first two of these forces both result from air pressure differences around an individual particle. Higher velocity winds are associated with lower air pressure, so where wind flow is accelerated over a particle lying on the surface there is also a decrease in pressure above the particle resulting in a lift force. Similarly, form drag results from the high wind pressure on the upwind side of the particle contrasting with the decreased pressure in the downwind region. These two pressure forces combine with the surface drag resulting directly from the shearing stress of the wind to shake the particle loose before spinning it up into the airstream.

The relationship between erosivity and entrainment can effectively be simplified to two parameters, critical wind shear (u_{*ct}) and particle diameter (d) (Bagnold 1941):

$$u_{*ct} = A\sqrt{\frac{(\sigma-\rho)}{\rho}\, g\cdot d}$$

where: σ = particle density, g = acceleration due to gravity, A = constant dependent upon the grain Reynolds number (≈ 0.1).

Generally, larger grains require a greater wind shear to dislodge them. However, as shown in Figure 1, this relationship is reversed for particles smaller than about 0.06 mm (dust-sized) where increased electrostatic and molecular cohesion require larger erosive forces for entrainment. Figure 1 also demonstrates that the grain sizes most susceptible to entrainment have diameters between 0.06 and 0.40 mm, sand-sized particles. It is this susceptibility of sand to entrainment that allows the accumulation of extensive dunefields (see SAND SEA AND DUNEFIELD) and SANDSHEETS in dryland regions.

A further process important in the entrainment of sand grains is the bombardment of the surface by grains that are already in transport. Once a few grains have been entrained by the wind they may be transported by the process of saltation,

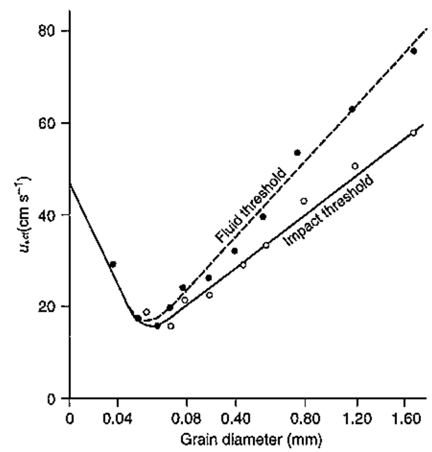

Figure 1 The relationship between particle size and the threshold shear velocity required for entrainment (after Chepil 1945)

Source: Thomas, D.S.G. (1997) *Arid Zone Geomorphology*, 2nd edition. © John Wiley & Sons Limited. Reproduced with premission bouncing along the surface with ballistic trajectories (see below). Each saltating grain gathers momentum from the wind and then imparts it to the sand surface on impact. This impact can 'splash-out' up to ten other grains that may also become entrained by the wind. A few saltating grains can quickly induce mass transport of sediment in a cascading system (Nickling 1988). Two thresholds of entrainment may therefore be identified: one (the fluid threshold) relates only to the drag and lift forces of the wind, the second (the impact threshold) is lower and combines wind forces with additional forces provided by impacting grains already in transport (see Figure 1). Once a sediment surface has begun to be eroded by wind forces at the fluid threshold, sediment transport is maintained at the lower impact threshold because energy is also available from the saltating grains. Wind shear stress may therefore reduce once entrainment has begun, but sediment transport will continue until the wind drops below the new impact threshold.

Transport mechanisms

The grain size of entrained particles also determines the mode of transport undertaken. Although dust-sized particles are not the easiest to entrain, they have very low settling velocities in comparison to potential wind lift and turbulent velocities, so can be transported in *suspension*. Particles suspended in the atmosphere may be held aloft for several days and hence travel long distances. An example of this is the deposition of Saharan dust in the south-eastern USA (Prospero 1999). Often this transport in suspension is in barely visible dust haze, but sometimes there are more concentrated dust plumes which are clearly seen on spectacular satellite images, and still less frequently suspended aeolian material is concentrated as dust storms which can lead to 'blackout' conditions.

Particles up to about 1.0 mm are commonly transported in SALTATION. Figure 2 shows the typical trajectory of saltating particles with a progressively increasing forward velocity from entrainment to impact as the particle draws momentum from the wind.

The actual trajectory of a particle depends on the height of its bounce. Wind velocity increases at a logarithmic rate away from the surface and so a particle that bounces higher into the wind will be able to draw greater momentum from it and so travel further and faster. The length of jump is thought to be about 12–15 times the height of bounce, or further if the particle spins

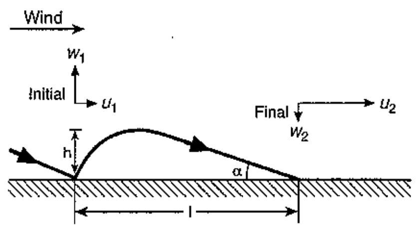

Figure 2 The ballistic trajectory of a saltating sand grain. w and u represent vertical and horizontal velocities, respectively (after Bagnold 1941)

Source: Thomas, D.S.G. (1997) *Arid Zone Geomorphology*, 2nd edition. © John Wiley & Sons Limited. Reproduced with premission

and induces an additional lift force (called the Magnus effect, White and Schultz 1977). The height of the saltation layer is dependent both on the wind velocity and also on the hardness of the surface over which the particles are saltating. Sand that is saltating over a rock or pebbly surface loses much less of its momentum on impact and so tends to bounce higher, reaching up to 3.0 m. An average saltation height, however, is about 0.2 m. The amount of sand in transport declines exponentially with height and so up to 80 per cent of all saltation activity takes place within 0.02 m of the surface (Butterfield 1991).

Grains which are ejected into the airflow as a result of the impact of a saltating grain may also enter the saltation system. However, some of these ejected grains may not have sufficient velocity fully to enter saltation and hence take only a single jump in a downwind direction. This process is termed *reptation* and, whilst much further research is required into its operation, it may be very significant in near-surface aeolian transport (Anderson *et al.* 1991).

The final mode of sand transport is *creep* and this describes the downwind rolling of larger sand particles (usually >0.5 mm). Such a process results both from the drag of wind on the surface of the particles and also the high velocity impact of saltating grains. It is thought that the process of creep may account for up to one-quarter of the bedload (saltation plus creep) transport rate.

Sand flux

The mass flux of sand (q) transported during an erosion event is often calculated as a cubic function of wind shear velocity (u_*). Most relationships

are derived from theoretical analyses or wind tunnel experiments and a popular expression is that of Lettau and Lettau (1978):

$$q = C\left(\frac{d}{D}\right)^{0.5}(u_* - u_{*ct})u_*\frac{2\rho}{g}$$

where: C = constant (4.2), d = grain diameter, D = standard grain diameter (0.25 mm), u_{*ct} = shear velocity threshold of grain entrainment, ρ = air density, g = acceleration due to gravity.

There has been little empirical testing of relationships like the one above and that which has been accomplished shows considerable variation between observed and predicted rates. Such variation is to be expected when the complex nature of the saltation system is considered and the fact that the predictive expressions available rarely account for variations in terrain, vegetation, surface moisture or wind turbulence. Furthermore, the accurate measurement of sand flux in the natural environment is very difficult, with the published efficiencies of sand traps varying between 20 and 70 per cent (Jones and Willetts 1979).

Deposition

Just as material is entrained when shear stress overcomes inhibiting forces, so material is deposited when shear stress is no longer greater than these forces. This manifests itself both at the large scale where, if regional wind patterns lead to a decrease of wind speed, SANDSHEETs or dunefields (see SAND SEA AND DUNEFIELD) are formed, but also at the much smaller scale where surface irregularities may be responsible for the deposition of sand patches. Finer-grained material carried in suspension is often deposited as LOESS.

Although a lack of vegetation is usually important in the aeolian entrainment of material, paradoxically its presence can be important in trapping dust and sand in depositional features. Dust is only deposited as loess where it is prevented from re-entrainment, often by vegetation, and vegetation can also be important in stabilizing coastal dune ridges.

References

Anderson, R.S., Sørensen, M. and Willetts, B.B. (1991) A review of recent progress in our understanding of aeolian sediment transport, *Acta Mechanica Supplement* 1, 1–19.

Bagnold, R.A. (1941) *The Physics of Blown Sand and Desert Dunes*, London: Methuen.

Bullard, J.E. and Livingstone, I. (2002) Interactions between aeolian and fluvial systems in dryland environments, *Area* 34, 8–16.

Butterfield, G.R. (1991) Grain transport rates in steady and unsteady turbulent airflows, *Acta Mechanica Supplement* 1, 97–122.

Chepil, W.S. (1945) Dynamics of wind erosion: 1. Nature of movement of soil by wind, *Soil Science* 60, 305–320.

Jones, J.R. and Willetts, B.B. (1979) Errors in measuring uniform aeolian sandflow by means of an adjustable trap, *Sedimentology* 26, 463–468.

Kocurek, G. and Lancaster, N. (1999) Aeolian system sediment states: theory and Mojave Desert Kelso Dune Field example, *Sedimentology* 46, 505–515.

Lettau, K. and Lettau, H.H. (1978) Experimental and micrometeorological field studies on dune migration, in H.H. Lettau and K. Lettau (eds) *Exploring the World's Driest Climates*, University of Wisconsin-Madison, Institute for Environmental Studies, Report 101, 110–147.

Nickling, W.G. (1988) The initiation of particle movement by wind, *Sedimentology* 35, 499–511.

Prospero, J.M. (1999) Long-term measurements of the transport of African mineral dust to the south-eastern United States: implications for regional air quality, *Journal of Geophysical Research* 104, 15,917–15,927.

White, B.R. and Schultz, J.C. (1977) Magnus effect in saltation, *Journal of Fluid Mechanics* 81, 497–512.

Further reading

Livingstone, I. and Warren, A. (1996) *Aeolian Geomorphology*, London: Longman.

Nickling, W.G. (1994) Aeolian sediment transport and deposition, in K. Pye (ed.) *Sediment Transport and Depositional Processes*, 293–350, Oxford: Blackwell Scientific.

Wiggs, G.F.S. (1997) Sediment mobilisation by the wind, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology*, 2nd edition, 351–372, London: Wiley.

GILES F.S. WIGGS AND IAN LIVINGSTONE

AEOLIANITE

Aeolianite is a cemented sandstone that has formed as a result of the processes of entrainment, transportation and deposition by wind. This rock type has various names including eolianite (US), miliolite (India and the Middle East), dunerock (South Africa), kurkar (Israel) and grès dunnaire (Mediterranean). Aeolianites of Quaternary age are most commonly found between 20° and 40° either side of the equator although examples have been found as far north as about 60°. Most aeolianites in the geological record are Quaternary in age and these tend to range between about 0.5 m to about 100 m in thickness.

Most aeolianites are rich in carbonate although silica-dominated forms also exist. Carbonate-rich aeolianites are largely associated with coastal sources of sediment and are thus close to modern or palaeo-shorelines. Semi-arid to sub-humid tropical shorelines are the most suitable locations for aeolianites as the oceans are productive in the formation of shelly biogenic grains or ooliths; strong onshore currents breakdown and move the sediment onshore; and climatic conditions are conducive for subsequent DIAGENESIS. In arid environments where there are strong onshore winds and a high sediment supply, dunes may be transported several hundred kilometres inland.

Along coastlines aeolianites often form elongate shore parallel or oblique bodies deposited as transverse ridges. The dunes are often stacked up against one another and may coalesce. The sediment size of aeolianites is typically coarse silt to sand-sized. Clays are rare because those that are entrained by the wind tend to be removed by suspension. Insufficient wind energy to transport grain sizes coarser than 2 mm generally limits the upper size of the clasts.

Aeolianites have distinctive bedding structures such as cross-bedding and laminations, which represent the progradation and growth of the dunes. Steeply dipping units up to 30°–34° reflect the former dune slipface. As a result of erosion the palaeodune bedforms are commonly lost and the dune type and direction of sand movement has to be deduced largely from the internal structures.

Lithification under freshwater vadose, mixed and/or phreatic conditions may occur. The main diagenetic processes result in alteration of unstable aragonite and high-Mg calcite clasts to low-Mg calcite clasts and cement. The balance between dissolution of carbonate grains by leaching and the production of cement is the prime control on the degree of dune induration. The main sources for the cement come from biogenic skeletal remains (e.g. molluscs, foraminifera, echinoderms, algae and coral fragments), ooliths, biota, sea spray, dust, bedrock and ground water.

Alteration of aeolianites by diagenetic processes in the vadose environment is the most common, occurring in three ways: (1) by loss of Mg^{2+} from the crystal lattices of high-Mg calcite; (2) by dissolution of aragonite, the loss of some strontium and reprecipitation as low-Mg calcite; and (3) by calcitization of aragonite *in situ*. A wide range of controls results in significant variability in aeolianite diagenesis in terms of both causal factors and diagenetic product (Gardner and McLaren 1994). Such controls include climate, sea level and time at the macro-scale; sea spray, plants and texture at the meso-scale and at the micro-scale the amount, rate of movement and chemistry of pore waters.

Major unconformities in aeolianites are often marked by PALAEOSOLs that develop as a result of solution and weathering. Commonly these soils are *terra rossa* and red latosols that have developed *in situ* but may contain inputs from windblown dust. Weathering may subsequently result in a solution of carbonate products and karstification. In semi-arid environments surface crusts and thin laminar CALCRETEs often form as a result of solution and rapid reprecipitation.

Radiometric dating of aeolianites is notoriously difficult. The effects of diagenesis mean that there are chances of contamination from secondary calcite. Occasionally unaltered shells are found which have allowed radiocarbon dating (e.g. McLaren and Gardner 2000). In addition, uranium series dating, amino acid racemization, luminescence and electron spin resonance dating have been used with varying success (see Brooke 2001).

Lithification increases the aeolianites' resistance to erosion and enhances their preservation potential in the geological record. The cement types (such as meniscus, rim, pore filling and needle fibre), amount and distribution, along with geochemistry, can aid interpretations concerning palaeoenvironments such as identifying palaeowater tables, palaeo-erosion surfaces or degree of exposure to marine environments.

References

Brooke, B. (2001) The distribution of carbonate eolianite, *Earth-Science Reviews* 55, 135–164.

Gardner, R.A.M. and McLaren, S. (1994) Variability in early vadose carbonate diagenesis in sandstones, *Earth-Science Reviews* 36, 27–45.

McLaren, S. and Gardner, R.A.M. (2000) New radiocarbon dates from a Holocene aeolianite, Isla Cancun, Quintana Roo, Mexico, *Holocene* 10, 757–761.

SEE ALSO: karst

SUE McLAREN

AGGRADATION

The long-term accumulation of sediment in a channel and the readjustment of the stream profile where there is a vertical growth of the land surface in response. Some possible agents of this process are running water, waves, glaciers and wind. Aggradation can occur at a variety of spatial scales and temporal scales (gradual or PUNCTUATED AGGRADATION), and may take place under constrained or unconstrained conditions. As aggradation is a long-term process, short-term fluctuations in sediment transport have no relevance.

Further reading

Easterbrook, D.J. (1999) *Surface Processes and Landforms*, 2nd edition, Upper Saddle River, NJ: Prentice Hall.

STEVE WARD

ALAS

The periglacial landform term alas, which is of Yakutian origin, was first introduced by the Russian worker P.A. Soloviev in the 1960s. It refers to the substantial circular and oval depressions with steep sides and flat floors, sometimes occupied with lakes, which characterize the geomorphology of the higher river terraces in central Yakutia (65°N, 125°E). Alases typically have diameters from 0.1 km up to 15 km and depths in the 3–40 m range. In morphological expression a tract of alas depressions on a river terrace surface is not dissimilar to a suite of KETTLE AND KETTLE HOLE forming a pitted glacial outwash plain. Genetically, an alas is a type of THERMOKARST feature, i.e. a subsidence landform arising from the degradation and settlement of ice-rich frozen ground. An essential prerequisite for alas development is terrain with a high ground ice content. The natural vegetation cover is taiga (coniferous forest) although the alas floors are often grass covered.

Coalescence processes amongst individual alas depressions can lead to the development of alas valleys. These valleys are characterized by a variable width with an alternation of narrow sections marking the former location of watersheds and wide sections together with branches with no outlets. The longitudinal profiles are not necessarily graded, reflecting the fact that they are thermokarst landforms rather than normal river cut features. In the Yakutian lowlands drained by the Lena River the spread of alas-related depressions has affected some 40 per cent of the higher river terrace surfaces. Surprisingly, the alas valleys form pockets of cultivated land where hardy strains of some grains along with some root crops are produced during the relatively warm summers.

Some prominence has been given in periglacial texts to a hypothetical reconstructed sequence of alas development (Soloviev 1973) although it needs to be emphasized that its applicability outside Yakutia has yet to be established. An important factor is that Yakutia is unglaciated yet nevertheless sustained permafrost throughout much of the Quaternary. Accordingly its ground ice history is complex with, for example, massive syngenetic ice wedges attaining sizes well in excess of those formed epigenetically.

The initial stage in alas development is a disturbance to the ground surface thermal regime's equilibrium state, such as can arise from the destruction of the natural vegetation as the result of climatic change, a forest fire or human activities. The upset thermal balance invariably leads to the degradation of the ice wedge tops beneath their surface polygonal troughs and the resultant growth of an enhanced hummocky surface morphology. Once ponded water accumulates between the mounds, further ice wedge decay is inevitable as in summer the water quickly warms and heat is transferred to the ice beneath. Thaw settlement in conjunction with progressive amalgamation of the ponds accelerates the melting process and leads to the creation of a thermokarst lake at the bottom of a major flat-bottomed depression. With time stability may be attained and lakes occupying old alas depressions may disappear through either sedimentation or drainage. Either process causes the sub-lake taliks to shrink leading to the growth of one or more PINGOs beneath the now dry hollow floor.

Reference

Soloviev, P.A. (1973) Thermokarst phenomena and landforms due to frost heaving in central Yakutia, *Builetyn perglacjalny* 23, 135–155.

SEE ALSO: ice wedge and related structures; permafrost; pingo; thermokarst

PETER WORSLEY

ALIGNED DRAINAGE

A parallelism of drainage lines. In some cases parallel or aligned drainage (rather than more normal dendritic patterns) covers great areas. As W.L. Russell (1929: 249) wrote:

> One of the most remarkable features of the northwestern Great Plains is the well-defined northwest–southeast alignment of the valleys

and ridges. The prevailing direction of the valleys and ridges is nearly identical over such great areas that it is evident that the causes or forces which produced the parallelism must have operated on a grand scale.

Using available maps, Russell showed that the alignment was developed in parts of western South Dakota, western Nebraska, western North Dakota, western Montana and eastern Wyoming. He suggested that this alignment was not caused by any structural control (though in other areas this is a perfectly valid hypothesis) but was associated with the former presence of sand dunes and associated interdunal channelling of erosion, particularly in the susceptible Pierre Shale. The aeolian hypothesis was endorsed by Flint (1955) who pointed out that alignment could be produced either by the former existence of linear dunes or by deflation of susceptible materials such as the Pierre Shale. Aligned drainage also occurs on the High Plains of Texas.

Similarly large expanses of aligned drainage occur in parts of Africa and are associated with the former greater extents of the Kalahari and Sahara deserts. In southern Angola, for example, even in areas where the current mean annual precipitation is as high as 1,200 mm, aligned stream channels (many of which are tens of km in length) run from east to west, as do the old dunes of the Mega-Kalahari (Thomas and Shaw 1991). In west Africa aligned drainage, related to the Pleistocene expansion of the Sahara, occurs as far south as southern Nigeria and Cameroon (Nichol 1998).

References

Flint, R.F. (1955) Pleistocene geology of eastern South Dakota, *US Geological Survey Professional Paper* 262.

Nichol, J.E. (1998) Quaternary climate and landscape development in west Africa: evidence from satellite images, *Zeitschrift für Geomorphologie NF* 42(3), 329–347.

Russell, W.L. (1929) Drainage alignment in the western Great Plains, *Journal of Geology* 37, 240–255.

Thomas, D.S.G. and Shaw, P.A. (1991) *The Kalahari Environment*, Cambridge: Cambridge University Press.

A.S. GOUDIE

ALLOMETRY

Allometry is the measurement of proportional changes in parts of an organism and correlated with variation in size of the total organism (Gould 1966). Church and Mark (1980) have provided the most comprehensive discussion of the geomorphological applications of allometry.

Allometric relations are usually described by power laws, such as $y = ax^b$ where x is an index of system scale, y is an attribute of the system, a is a constant and the exponent b is the ratio of x and y.

Four distinctions need to be made:

1 *Allometry and isometry* If x and y have the same dimensions, isometry obtains when $b = 1$. Under this condition, there is no change in the relative proportions of x and y with increasing scale and the system is described as self-similar. When $b \neq 1$, the relation is allometric, implying a scale-related distortion of geometry. For example, in Bull's (1964) analyses of the areas of alluvial fans compared with their contributing drainage areas, $b = 0.9$ and allometry obtains. This is an indication that larger drainage basins have a relatively greater tendency to store sediment than smaller basins.

2 *Negative and positive allometry* If $b > 1$, the relation is positively allometric; if $b < 1$, the relation is negatively allometric. However, care must be taken to check that the dimensions of x and y are the same. If, for example, y is a length (L) and x is an area (L^2), then a value of b of 0.5 would indicate isometry and values of $b >$ or < 0.5 would indicate positive and negative allometry respectively.

3 *Dynamic and static allometry* In biology it is relatively easy to compare organisms at various stages of growth (dynamic allometry). Typically, landforms are compared at one moment in time with little control over their absolute ages (static allometry). There are serious limitations to static allometry, not least of which is the spatial heterogeneity of geological materials and the difficulty of defining drainage basins with similar growth histories.

4 *Simple and compound allometry* If the b value of an allometric relation changes as system scale changes compound allometry obtains. There is increasing evidence that compound allometry results from dominant process change between slope-dominated and channel-dominated basins.

References

Bull, W.B. (1964) Geomorphology of segmented alluvial fans in western Fresno County, California, *US Geological Survey Professional Paper* 352-E, 89–129.

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

Gould, S.J. (1966) Allometry and size in ontogeny and phylogeny, *Cambridge Philosophical Society Biological Reviews* 41, 587–640.

SEE ALSO: fractals

OLAV SLAYMAKER

ALLUVIAL FAN

Alluvial fans are depositional landforms created where steep high-power channels enter a zone of reduced STREAM POWER. Typically they range in scale from axial lengths of tens of metres to tens of kilometres. They are usually cone-shaped forms with surface slopes radiating away from an apex, located at the point where the feeder channel enters the fan. This form can be modified by the presence of confining neighbouring fans or valley walls. In addition, the burial of the fan apex area can cause backfilling into the mountain catchment. Alluvial fans are subaerial features, however if they extend into water they are known as fan deltas.

Many of the classic studies of alluvial fans, which established the basic properties, were carried out in the basin-and-range terrain of the deserts of the American south-west (Blackwelder 1928; Blissenbach 1954; Hooke 1967), culminating in Bull's (1977) review paper. Since then there have been many studies of other (mostly) dry-region fans, with emphasis on relations between sedimentary processes and morphology (Wells and Harvey 1987; Blair and McPherson 1994) and on fan dynamics (Harvey 1997), in addition to studies of fans in humid regions (see Rachocki and Church 1990).

Fan occurrence

Alluvial fans occur in two characteristic situations: at mountain fronts and at tributary junctions. In both cases, high sediment loads encounter zones of reduced stream power, with accommodation space for deposition. These conditions are controlled by long-term landform evolution, including the tectonic setting and erosional history. Mountain fronts may be fault-controlled or erosional, in which case the fans may bury an older PEDIMENT surface. Tributary-junction settings are controlled by the long-term dissectional history. A common fan setting occurs in glaciated mountain terrain, where steep tributary valleys join wide formerly glaciated valleys.

Much of the literature emphasizes the importance of alluvial fans in desert mountain areas. In such areas, FLASH FLOODs transport abundant coarse sediment, and the depositional setting created by regional tectonics may be enhanced by the tendency for desert floods to lose power downstream. However, neither active tectonics nor aridity are prerequisites for fan formation. Fans can occur in mountain areas in all climatic settings and in tectonically stable areas, provided there is juxtaposition of high coarse-sediment transport and a sudden downstream loss in transporting power. But, as outlined below, as fan morphology tends to respond to climatically controlled water and sediment supply, climatic change can induce a change in fan processes and morphology.

Fan processes

Processes on fans include four groups. Primary processes deliver sediment to the fan, principally by DEBRIS FLOWs, or fluvial processes (by channelized and/or sheet flows). These processes are expressed by the sediments comprising the fan and by the surface morphology. Debris flows are massive, usually matrix-supported coarse sediments with clasts up to boulder size. Depositional features may include lobate and levee forms. Fluvial sediments in fan environments are usually moderately sorted gravels and cobbles, stratified or lensed in channelized or sheet bodies. Depositional features may include a range of channel forms or shallow bar and swale topography.

Fans have been classified on the basis of the primary processes into debris-flow and fluvially dominant fans. These processes are catchment controlled and depend on the water:sediment mix fed to the fan during flood events. Debris flows operate as sediment-rich flows, but under conditions of greater dilution become transitional or HYPERCONCENTRATED FLOWs then fluvial flows. Debris flows are most common where sediment concentrations are high, e.g. from small, steep catchments (Kostaschuk *et al.* 1986). Fluvial processes are more common from large, less steep catchments. The old-fashioned distinction between 'dry' and 'wet' fans, interpreted on the basis of climate, is outmoded – the primary processes are controlled mainly by catchment characteristics.

Secondary processes rework the sediment on the fan by fluvial, or in arid areas by AEOLIAN

PROCESSES. Third, stabilization processes involve surface modification by soil formation and vegetation colonization. Such processes may influence the hydrology of the fan surface, but are important in fan studies as they allow the relative ages of fan segments to be assessed (see McFadden *et al.* 1989). In arid and semi-arid areas such processes include surface modification by desert

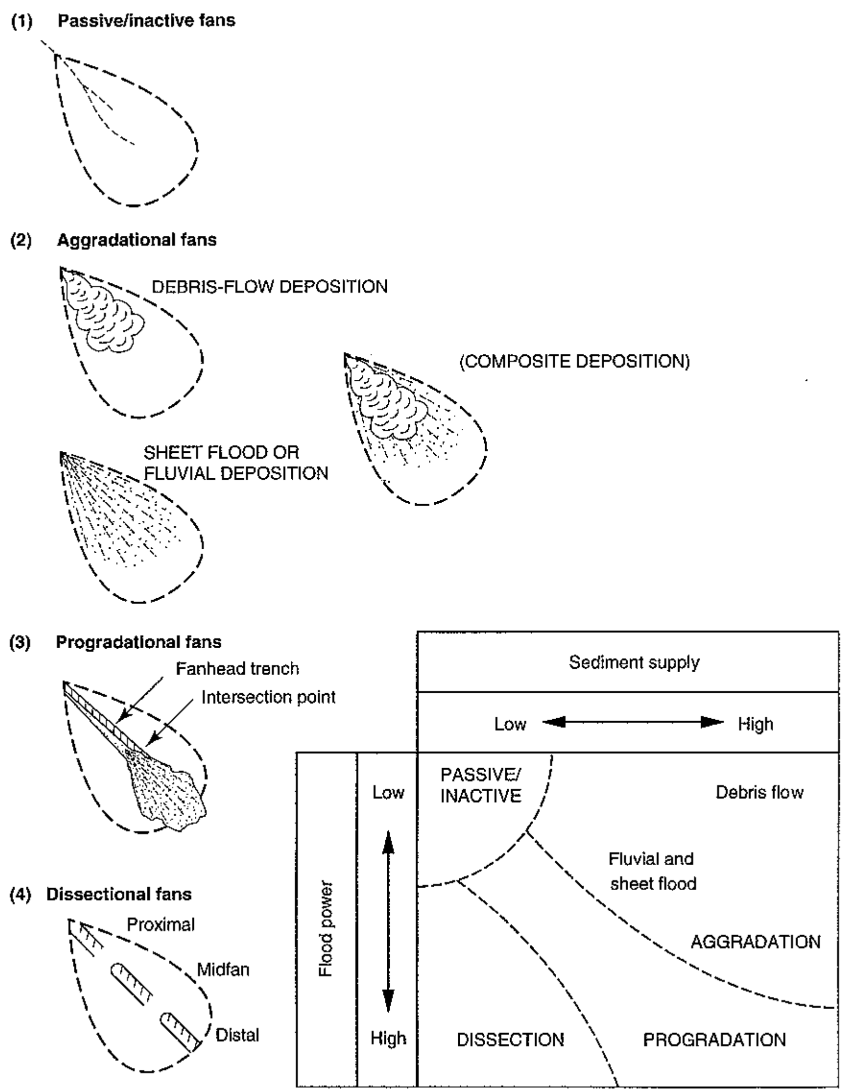

Figure 3 Alluvial fan styles: response to flood power and sediment supply (modified from Harvey 2002c)

pavement (see STONE PAVEMENT) formation and DESERT VARNISH development, and pedogenic processes leading for example to carbonate accumulation and CALCRETE formation. In humid areas lichen colonization and colonization by higher plants may be important as well as soil formation. Finally, dissection processes may erode the fan surface. Dissection may simply increase with fan surface age, or be accelerated by climatic or base-level change.

Fan morphology

Within the context of the topographic setting, fan morphology reflects fan processes and evolution. The relationships between erosion and deposition on the fan can be described as fan style (Figure 3), which in turn depends on the relationship between flood power and sediment supply. Under conditions of low power and little sediment supply the fan may be inactive. Under conditions of excess sediment supply the fan will aggrade, by debris-flow or fluvial processes dependent on the water:sediment mix fed to the fan. Such AGGRADATION will occur from the fan apex downfan. Commonly, both power and sediment supply are moderate. The feeder channel incises into the fan surface to form a fanhead trench, which emerges onto the fan surface at a midfan intersection point (Plate 1), beyond which deposition occurs. Such fans are described as 'telescopic' and may extend by progradation. A zone of coalescent deposition from adjacent prograding fans is known as a BAJADA. If power

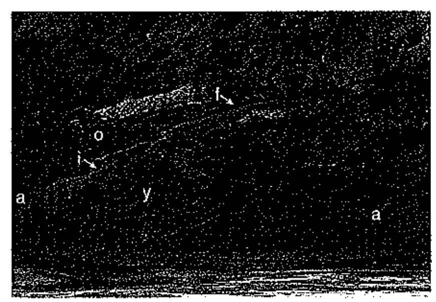

Plate 1 Characteristic alluvial fan morphology: Death Valley, California (photo: A.M. Harvey)

Notes: f = fanhead trench; i = intersection point; o = older fan surfaces; y = younger fan surfaces; a = active depositional segment

is excessive, either through high runoff or sediment starvation, erosion may dominate. Erosion may be concentrated within the fanhead area, in midfan, or in the case of base-level induced erosion, at the fan toe.

On many fans, BASE LEVELS are stable, at least over moderate timescales, and fan processes are primarily proximally controlled – by water and sediment supply from the catchment. A climatic (or other environmental change, e.g. related to human activity) causing changes in water and sediment supply may result in a change in fan style towards greater erosion or deposition.

Two aspects of fan morphometry have been demonstrated to reflect fan context, processes and evolution. General relationships of fan area and fan gradient to drainage areas have the forms:

$$A_f = p\, A_d^{q} \quad (1)$$
$$G_f = a\, A_d^{-b} \quad (2)$$

(where A is area, G is gradient, f of the fan, d of the drainage basin, pqab are constants). For the fan-area relationship, exponent q generally ranges between 0.7 and 1.1, and the value of the constant p reflects fan age, degree of confinement, basin area, geology and climate. For the fan-gradient relationship exponent b generally ranges between −0.15 and −0.35, and the value of constant a primarily reflects sedimentary processes (Harvey 1997). Debris-flow fans are steeper than fluvial fans. Fan-surface and fan-channel profile relationships (Figure 4) reflect erosion and deposition histories, and the interaction between proximal climate- and sediment-led controls and distal base-level controls.

Fan dynamics

Three sets of factors affect the geomorphology of alluvial fans: (1) context and locational factors, particularly tectonics and geomorphic history; (2) water and sediment delivery to the fan, controlled in the context of catchment geology size and relief, largely by climatic factors; (3) factors affecting the fan environment itself, especially base level.

Interactions between tectonic, climatic and base-level factors form a major thrust of alluvial fan research. Tectonics and gross geomorphology may control the fan setting, but the consensus is that, at least for Quaternary fans, climate appears to have the primary role in causing changes in fan

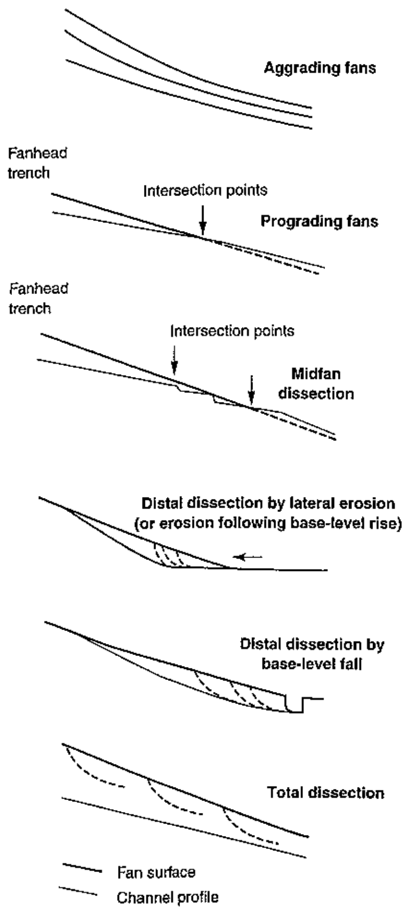

Figure 4 Fan surface and fan channel relationships

behaviour and dynamics (Frostick and Reid 1989; Ritter *et al.* 1995), modified by base-level conditions (Harvey 2002a).

Significant changes in fan processes in response to Quaternary climatic changes have been identified in many areas, including dry regions (e.g. Wells *et al.* 1987; Bull 1991), and humid temperate regions, especially in a PARAGLACIAL context in mountain areas glaciated during the Pleistocene (Ryder 1971).

Alluvial fans are important features within mountain fluvial systems. They act as sediment stores, modifying the transmission of coarse sediments through the fluvial system. They have a profound effect on the buffering/coupling relationships of fluvial systems (Harvey 1997, 2002b). Similarly they preserve a sensitive sedimentary record of environmental change within the mountain source areas.

References

Blackwelder, E. (1928) Mudflow as a geologic agent in semi-arid mountains, *Geological Society of America Bulletin* 39, 465–484.

Blair, T.C. and McPherson, J.G. (1994) Alluvial fan processes and forms, in A.D. Abrahams and A.J. Parsons (eds) *Geomorphology of Desert Environments*, 354–402, London: Chapman and Hall.

Blissenbach, E. (1954) Geology of alluvial fans in semi-arid regions, *Geological Society of America Bulletin* 65, 175–190.

Bull, W.B. (1977) The alluvial fan environment, *Progress in Physical Geography* 1, 222–270.

——(1991) *Geomorphic Responses to Climatic Change*, Oxford: Oxford University Press.

Frostick, L.E. and Reid, I. (1989) Climatic versus tectonic controls of fan sequences: lessons from the Dead Sea, Israel, *Journal of the Geological Society, London* 146, 527–538.

Harvey, A.M. (1997) The role of alluvial fans in arid zone fluvial systems, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology: Process, Form and Change in Drylands*, 2nd edition, 231–259, Chichester: Wiley.

——(2002a) The role of base-level change in the dissection of alluvial fans: case studies from southeast Spain and Nevada, *Geomorphology* 45, 67–87.

——(2002b) Effective timescales of coupling within fluvial systems, *Geomorphology* 44, 175–201.

——(2002c) Factors influencing the geomorphology of dry-region alluvial fans: a review, in A. Perez-Gonzalez, J. Vegas and M.J. Machado (eds) *Aportaciones a la Geomorfologia de Espana en el Inicio del Tercer Mileno*, 59–75, Madrid: Instituto Geologico y Minero de Espana.

Hooke, R. le B. (1967) Processes on arid region alluvial fans, *Journal of Geology* 75, 438–460.

Kostaschuk, R.A., MacDonald, G.M. and Putnam, P.E. (1986) Depositional processes and alluvial fan – drainage basin morphometric relationships near Banff, Alberta, Canada, *Earth Surface Processes and Landforms* 11, 471–484.

McFadden, L.D., Ritter, J.B. and Wells, S.G. (1989) Use of multiparameter relative-age methods for age estimation and correlation of alluvial fan surfaces on a desert piedmont, eastern Mojave Desert, California, *Quaternary Research* 32, 276–290.

Rachocki, A.H. and Church, M. (eds) (1990) *Alluvial Fans: A Field Approach*, Chichester: Wiley.

Ritter, J.B., Miller, J.R., Enzel, Y. and Wells, S.G. (1995) Reconciling the roles of tectonism and climate in Quaternary alluvial fan evolution, *Geology* 23, 245–248.

Ryder, J.N. (1971) The stratigraphy and morphology of paraglacial alluvial fans in south central British Columbia, *Canadian Journal of Earth Sciences* 8, 279–298.

Wells, S.G. and Harvey, A.M. (1987) Sedimentologic and geomorphic variations in storm generated alluvial fans, Howgill Fells, northwest England, *Geological Society of America Bulletin* 98, 182–198.

Wells, S.G., McFadden, L.D. and Dohrenwend, J.C. (1987) Influence of late Quaternary climatic change on geomorphic and pedogenic processes on a desert piedmont, eastern Mojave Desert, California, *Quaternary Research* 27, 130–146.

ADRIAN HARVEY

ALLUVIUM

Alluvium (neuter of the Latin adjective *alluvius*, meaning washed against) is the term used for the sediments that are deposited by flowing water in river valleys and deltas. Alluvial sediment originates ultimately from the breakdown (weathering) of pre-existing rocks on land. This sediment is then transported downslope by mass wasting, overland water flow, river flow and floodplain flow, and deposited in areas of the river valley where the water flow decelerated. Alluvium is deposited in distinctive landforms (e.g. channel bars, channel fills, levees, crevasse splays, flood basins, fans and deltas). Alluvial sediments are normally stratified gravels, sands, silts and clays, and the texture and stratification of the sediments are determined by the associated landform and the mode of deposition and subsequent erosion. Alluvium has been deposited on the Earth's surface for as long as rivers have existed. The sedimentary characteristics of modern alluvium are used to interpret the origin of ancient alluvium. Alluvium of modern rivers and floodplains commonly is fertile agriculturally, and a source for ground water, sand and gravel. Ancient alluvium commonly contains economically important resources such as water, gas, oil, coal and placer minerals. Reviews of the origin and nature of alluvial deposits are given by Bridge (2002), Carling and Dawson (1996) and Miall (1996).

Nature of transport and deposition of alluvium

Water flow in alluvial river channels and floodplains is turbulent. Turbulent water flow results in relatively coarse sediment (sand and gravel) being transported near the sediment bed as bedload, and finer-grained sediment (sand, silt and clay) being transported within the flow as suspended load. Depending on the sediment transport rate, the bed is normally moulded into various types of bedform, such as ripples, dunes and antidunes. Near-plane beds also occur in sands when the sediment transport rate is relatively high and in gravels at low sediment transport rates. The geometry of ripples is dependent on bed-sediment size. The geometry of dunes and antidunes is related to flow depth. Bars are larger bedforms that occur in channels, and their geometry is controlled mainly by channel width. Ripples, dunes and antidunes may be superimposed on bars.

Deposition of alluvium occurs mainly due to spatial (but also temporal) decrease in water flow velocity (actually bed shear stress) and sediment transport rate. This deposition occurs over a large range of spatial scales in areas of flow expansion and deceleration such as the lee sides of bedforms, at the edge of channels as water moves onto the floodplain, abandoned channels, zones of tectonic subsidence, and where water flows into lakes and the sea (forming deltas). Most deposition occurs during floods when flow velocity, bed shear stress and sediment transport rate are large. If sediment transport rate is large, spatial decrease in sediment transport rate will cause relatively high deposition rate. However, much deposited sediment is subsequently re-eroded during the same or subsequent floods. The coarsest sediments are deposited from bedload in places where bed shear stress is large (typically channels), whereas the finest grained sediments are deposited from suspended load (typically in abandoned channels, flood basins and lakes). Intermediate sediment sizes are deposited from bedload and suspended load. It is common for the size of channel-bed sediment to decrease down valley, primarily due to downstream decrease in channel slope and bed shear stress. It is also common for bed-sediment size to decrease laterally from the channel to the distal edge of the floodplain, also due to decreasing bed shear stress.

Alluvial landforms

Alluvial river channels contain various types of bars, the geometry and evolution of which determine the plan form of the channel. Simple (unit) bars occur in all alluvial channels. In meandering channels, the unit bars combine to give compound point bars on the inside of channel bends. In braided channels, the unit bars combine to give mid-channel compound bars (braid bars) in addition to point bars. As the supply of water and/or sediment increases, alluvial channels change from meandering to braided. Straight channels are rare and occur when the stream is not powerful enough to erode its banks. Channels change position by bank erosion and bar deposition, or by channel diversions. Channels can be diverted within their channel belts by cutoff, and channel belts can be diverted to different positions within their floodplains (avulsion). Channels abandoned by cutoff or avulsion become blocked with bars and eventually become elongate lakes.

Floodplains are the areas adjacent to channels that are inundated with water during seasonal floods. LEVEES are wedge-shaped accumulations of sediment that form floodplain ridges adjacent to channels. Crevasse channels cut levees in places and pass downstream into lobate sediment accumulations called crevasse splays. Some levees are composed of laterally adjacent crevasse splays. Crevasse splays are fan shaped in plan and contain a system of distributive and or anastomosing channels. The active and abandoned channels, levees and crevasse splays constitute the alluvial ridge that stands above the adjacent flood basin. The alluvial ridge exists because deposition rate is greatest in and around the main channel. The flood basin contains floodplain channels, both ephemeral and permanent lakes, and abandoned channel belts (alluvial ridges).

Alluvial fans and deltas are areas of alluvial deposition that are distinctive because of their plan shapes and distributive and/or anastomosing channels bordered by floodplains. Deltas build into standing bodies of water. If fans build into standing bodies of water, they are referred to as fan deltas. The term terminal fan has been used for fans in arid areas where water flow percolates into the ground before reaching beyond the fan margins. Alluvial fans occur in all climates where a confined channel passes from an area of high slope to an unconfined area of lower slope. The abrupt change of slope results in a downstream decrease in bed shear stress and sediment transport rate, which leads to deposition. ALLUVIAL FANS commonly occur adjacent to fault scarps, and the preservation of fan deposits is enhanced by the subsidence of the hanging wall. Usually, one channel is active on a fan surface at any time, but avulsion is a common process and many wholly or partially abandoned channels occur on fan surfaces. Where fan surfaces are steep (and relatively coarse grained), sediment gravity flows are common depositional processes in addition to water flows.

A RIVER DELTA is a mound of sediment deposited where a river channel enters a body of water (such as a lake or sea) and supplies more sediment than can be carried away by currents in the water body. At the river mouth, the previously confined flow expands and decelerates, depositing its sediment load. The coarse bedload is deposited close to the mouth (as a mouth bar), whereas the finer sediment in suspension is carried further into the water body before being deposited. Currents in the body of water (perhaps associated with tides, wind waves, geostrophic flows or turbidity currents) may subsequently rework and move the deposited sediment. The morphology and sediments of deltas reflect the balance between these different stages of delta formation.

River terraces (see TERRACE, RIVER) are remnants of floodplains, fans or delta plains that have become elevated relative to the modern river and floodplain, as a result of widespread channel incision. Different episodes of incision and deposition can result in a series of terraces of different height, and valley fills with a complicated internal structure.

Alluvial deposits

Depending on the availability of different sediment sizes, channel deposits are usually mainly gravels and sands. Floodplain deposits are mainly sands, silts and clays. Different scales of stratification in alluvial deposits depend on the scale of topographic feature associated with the deposit: ripples form small-scale cross strata (set thickness <30 mm); dunes form medium-scale cross strata (set thickness 30 mm to metres); unit bars form simple sets of large-scale inclined strata (set thickness normally decimetres to metres); compound bars form compound sets of large-scale inclined strata (set thickness metres to tens of metres). Channel fills are composed of bar deposits overlain by lacustrine silts and clays. Channel belts are composed of superimposed bars and channel fills, and are commonly metres to tens of metres thick and hundreds to thousands of metres wide. Levees, crevasse splays and lacustrine deltas may be metres thick and hundreds to thousands of metres long and wide, composed mainly of sands and silts. Floodplain-channel fills typically are up to metres deep and tens to hundreds of metres across. Silty and clayey deposits of flood basins and lakes commonly occur in metre-thick sequences. Floodplain deposits are normally subjected to pedogenesis, and soil horizons are ubiquitous in alluvium. The nature and degree of soil development varies in time and space as a function of floodplain deposition rate, parent materials, groundwater composition, climate and vegetation.

The proportion of channel-belt deposits (coarse sediments) relative to floodplain deposits (fine sediments) in a valley fill depends on factors such as the frequency of channel-belt diversions (avulsions), the width of the channel belt relative to the floodplain width, the overall deposition rate, and tectonic subsidence or uplift within the valley. High proportions of channel deposits in valley fills typically occur on the upstream parts of alluvial fans where deposition rate and avulsion frequency are locally high, in parts of valleys that are narrow relative to the channel-belt width (e.g. incised valleys), and in areas of the valley where tectonic subsidence has attracted avulsing channel belts. Low proportions of channel deposits in valley fills occur typically where floodplain (valley) width is large relative to channel-belt width (e.g. on delta plains), and in tectonically uplifted parts of floodplains.

As avulsion frequency, relative widths of channel belts and floodplains, overall deposition rate, and subsidence or uplift rate are controlled by climate, eustatic sea-level change, and tectonism, the nature of the valley fill will also be controlled by these factors. Furthermore, spatial and temporal variations in the effects of climate, eustatic sea-level change and tectonism on deposition and erosion of alluvium result in spatial variations in its texture and internal structure. These spatial variations are commonly cyclic.

References

Bridge, J.S. (2002) *Rivers and Floodplains*, Oxford: Blackwells.

Carling, P.A. and Dawson, M.R. (eds) (1996) *Advances in Fluvial Dynamics and Stratigraphy*, Chichester: Wiley.

Miall, A.D. (1996) *The Geology of Fluvial Deposits*, New York: Springer-Verlag.

SEE ALSO: aggradation; anabranching and anastomosing river; antidune; avulsion; bank erosion; bar, river; bedform; bedload; braided river; channel, alluvial; current, downstream fining; erosion; dune, fluvial; flood; flood plain; point bar; suspended load

JOHN S. BRIDGE

AMPHITHEATRE

Some early studies of curved valley heads of non-glacial origin attributed them to erosion under arid climates, but they are also widespread in humid areas. They occur mainly where a valley extends headward through gently inclined sedimentary rocks, or through dissected volcanic domes, such as those of Hawaii. Unless angular morphology is maintained by strong rectangular fracturing, curved planimetry develops as a strong CAPROCK is undercut either by seepage or by mass failure.

An amphitheatre can be likened to an arch lying on its side, because lateral stresses hold blocks in place on the curved rock face. This is especially so where the dominant stresses in a rock mass are essentially horizontal, and keep the rock face in compression. Amphitheatres thus tend to be more stable than straight cliff lines. The development of the curvature seems to be linked to the three-dimensional distribution of stresses on the rock face. Experimental studies for open-cut mining show that slopes are most stable where the radius of curvature approximates the height on the back wall, but that stability decreases markedly as the radius of curvature increases to about four times the height. Similar relationships occur in many natural amphitheatres. South of Sydney, Australia, 90 per cent of amphitheatres have a radius-to-height ratio below 5:1, with approximately 20 per cent of them below 2:1. The dimensions of amphitheatres are far from random, and are indicative of an equilibrium between form and stress distribution.

Further reading

Laity, J. and Malin, M.C. (1985) Sapping processes and the development of theater-headed valley networks on the Colorado Plateau, *Geological Society of America Bulletin* 96, 203–217.

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

R.W. YOUNG

ANABRANCHING AND ANASTOMOSING RIVER

An *anabranching* alluvial river is a system of multiple channels characterized by vegetated or otherwise stable alluvial islands that divide flows at discharges up to bankfull (Plate 2). The islands may be developed from within-channel deposition, excised by channel AVULSION from extant floodplain, or formed by prograding distributary-channel accretion on splays or deltas. A specific

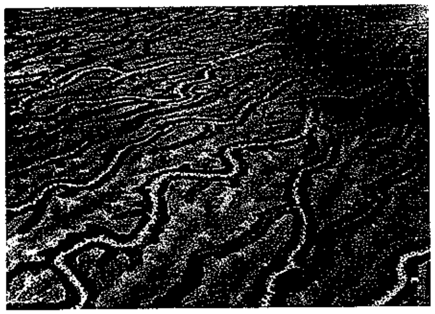

Plate 2 An aerial view of muddy anabranching channels at South Galway on Cooper Creek in western Queensland, Australia. Standing water is pale grey and the recently wetted channel boundary is darker and about 20–30 m wide. The islands were not quite over-topped by this bankfull flow

subset of distinctive low-energy anabranching systems associated with mostly fine-grained or organic sedimentation are defined as *anastomosing* rivers (Smith and Smith 1980; Knighton and Nanson 1993; Makaske 2001). Neither of these terms now applies to BRAIDED RIVERs where divided flow is strongly stage dependent around bars that are unconsolidated, ephemeral, poorly vegetated and overtopped at less than bankfull. However, some confusion remains because an individual low-flow channel in a braided system is sometimes referred to as an anabranch. The islands in an anabranching river are about the same elevation as the adjacent floodplain, persist for decades to centuries, have relatively resistant banks, and support mature vegetation.

Anabranching bedrock rivers can occur where the individual channels follow joint and fracture patterns. However, bankfull flow is unclearly defined making such rivers difficult to compare to their alluvial counterparts. Channels are often sediment free with pools, cataracts and waterfalls. Van Niekerk *et al.* (1999) found that bedrock anabranching channels on the Sabie River in South Africa have a significantly greater potential to transport sediment than do all the other channel types along that river. At present, relatively little is known about bedrock anabranching systems.

Anabranching is not a mutually exclusive category for it occurs in association with other patterns whereby individual anabranches braid, meander (see MEANDERING) or are straight, and it occupies a wide range of environments, from low to high energy, and in arctic, alpine, temperate, humid tropical and arid climatic settings. Anabranching rivers are more common than has been recognized previously; a total of more than 90 per cent by length of the alluvial reaches of the world's five largest rivers anabranch and it is a particularly widespread river pattern in inland Australia for both large and small rivers. In Europe many rivers used to anabranch but most of these have now been modified to provide more convenient single-thread systems in densely populated and heavily utilized valleys.

Determining the fundamental cause of anabranching remains ellusive but it is understood that in some cases, the advantage of anabranching over a single wide channel is that islands concentrate stream flow and maximize bed-sediment transport per unit of stream power, thereby maintaining equilibrium conditions. This occurs particularly where there is little or no opportunity to increase channel gradient (Nanson and Huang 1999) or where vegetation increases channel roughness (Tooth and Nanson 2000). In other words, some anabranching rivers appear to exhibit MAXIMUM FLOW EFFICIENCY and LEAST ACTION PRINCIPLE (Huang and Nanson 2000). However, there are also cases where anabranching is associated with non-equilibrium sediment transport and inefficient flow, exhibiting extensive overbank flooding, the dispersal of sediment over extensive floodplains (Plate 2), and rapid vertical accretion (Makaske 2001; Abbado *et al.* 2003). As with meandering and braiding rivers, it is apparent that anabranching systems can exhibit equilibrium or non-equilibrium behaviour.

Classification

Six types of anabranching river have been recognized by Nanson and Knighton (1996) on the basis of stream energy, sediment size and morphological characteristics: Types 1–3 are lower energy and Types 4–6 are higher energy systems. Figure 5 illustrates the planform expressions for various types of anabranching river. Type 1 consists of *cohesive sediment* rivers (commonly termed anastomosing rivers) with low w/d ratio channels that exhibit little or no lateral migration. Type 2 consists of *sand-dominated island forming* rivers and Type 3 consists of *mixed load laterally active*

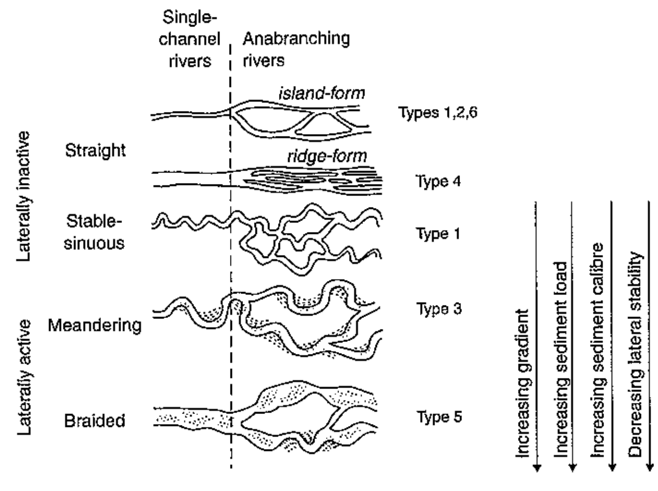

Figure 5 A classification of river channel patterns including single channel and anabranching planforms. Laterally inactive channels consist of straight and sinuous forms whereas laterally active channels consist of meandering and braided forms. The anabranching types are described in the text (after Nanson and Knighton 1996)

meandering rivers. Type 4 consists of *sand-dominated ridge form* rivers characterized by long, parallel channel-dividing ridges. Type 5 consists of *gravel-dominated laterally active* rivers that interface between meandering and braiding in mountainous regions. These have been described as wandering gravel-bed rivers (Church 1983). Type 6 consists of *gravel-dominated stable* rivers that occur as non-migrating channels in small, relatively steep basins.

Anastomosing rivers

Anastomosing rivers are an economically important subgroup of anabranching rivers and consequently have been studied in detail by sedimentologists because of their fine-grained nature and tendency to accumulate a substantial organic (coal) stratigraphy. Anastomosing commonly occurs in the lower fine-textured reaches of rivers, or in depositional basins, where vertical accretion can be rapid and hence their preservation potential is high. Crevasse splays and thick natural levees may be common. Makaske (2001) describes them as forming by avulsion and the islands as having flood basins, but these characteristics are not so apparent in some arid environments (Knighton and Nanson 1993). Modern examples were first described in detail in the alpine and humid environment of the Rocky Mountains of western Canada (e.g. Smith 1973; Smith and Smith 1980) but have subsequently been described in a wide variety of settings including arid environments (e.g. Knighton and Nanson 1993; Gibling *et al.* 1998; Makaske 2001).

Anastomosing river stratigraphy

In rapidly accreting humid settings, peats can accumulate in floodplain lakes and swamps to form coal, and sandy palaeochannels may act as reservoirs for hydrocarbons. However, not all anabranching rivers are rapidly vertically accreting and in arid environments they do not accumulate organics. Makaske (2001) found no standard sedimentary succession for anastomosing rivers, although he described them in three different settings and showed some common characteristics. The Columbia River is an example of the style of stratigraphy in a rapidly vertically accreting humid montane setting with organic-clastic accumulation (Figure 6). Such anastomosing rivers (and delta distributary

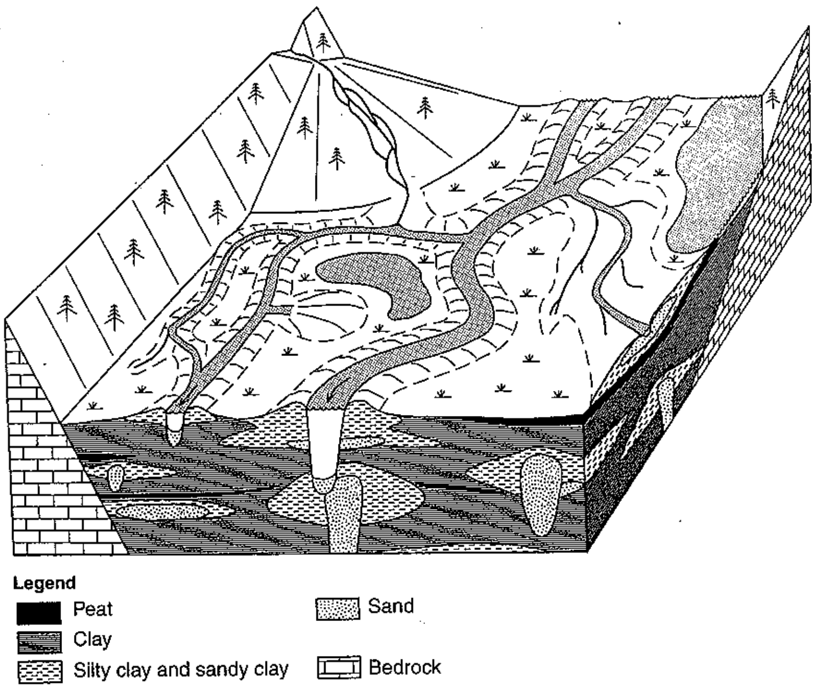

Figure 6 Textural facies model of the upper Columbia River (British Columbia, Canada), a rapidly aggrading anastomosing system in a temperate humid montane setting. Scale is approximately 2 km in width and alluvial thickness ~10 m (after Makaske 2001)

channels) tend to have fixed channels that aggrade with only limited lateral migration, thus generating ribbons or narrow sheets. Many of these deposits lie in rapidly subsiding settings, especially foreland and extensional basins characterized by large sediment flux and low gradients. The avulsion of a major channel into wetlands may generate a splay complex with suites of small, transient anastomosing channels, with the eventual establishment of a stable, single-channel course. In arid environments, alluvial and aeolian deposits can be juxtaposed, whereas in vertically accreting humid environments channel fills are flanked by silty levee deposits, lacustrine clay and coal. However, because it is difficult to show that the palaeochannels formed a synchronous anastomosing network at a single point in time (Makaske 2001), assessing the truly anabranching origin of such stratigraphies may be sometimes an educated guess.

Vegetation

Vegetation plays a major role in the development and maintenance of anabranching rivers. Indeed, it is very likely that truly anabranching rivers did not exist prior to the Devonian Period when the evolution of land plants and their associated role in the weathering of clays and the stabilization of the land surface became important. The establishment and maintenance of channels and islands with stable, often near vertical, banks means that the channels, instead of widening as a simple function of shear stress and limited alluvial strength, maintain narrow, deep and flow-efficient channels. Smith (1976) demonstrated the enormous increase in erosional resistance that plant roots can offer riverbanks. In some dryland rivers anabranching has been shown to increase in intensity below tributary junctions due to irrigation of the often dry channel floor and the greater flow and sediment-transport resistance

offered by trees growing on the channel bed (Tooth and Nanson 1999). Such anabranching, resulting from the progressive evolution of within channel bars to ridges, organizes the flow into well-defined multiple channels, narrower in total than the adjacent single-thread reaches (Wende and Nanson 1999; Tooth and Nanson 2000). In certain dryland environments where bankline trees are less dense, then cohesive mud plays an important role in producing stable multiple-channel systems (Gibling *et al.* 1998).

Conclusion

Anabranching characterizes a disparate group of alluvial systems from low-energy organic or fine sediment-textured, to high-energy gravel transporting rivers, and even occurs in bedrock systems. It is a widely represented – even the dominant – style along the world's largest alluvial rivers. Alluvial anabranching rivers can be equilibrium systems that maintain their sediment flux by confining bankfull flows, or non-equilibrium systems that very effectively distribute and deposit excess sediment over extensive depositional surfaces. Anabranching is commonly associated with flood-dominated flow regimes and well-vegetated, erosion-resistant banks. As such they sometimes exhibit mechanisms to block or constrict channels and induce channel avulsion. Some develop as erosional systems that scour channels into floodplains or jointed bedrock, while others build long-lived, stable islands or ridges within existing channels. On deltas they can build floodplains vertically around initially subaqueous channels. Anabranching rivers are commonly laterally stable but individual channels can meander, braid or be straight, and as such they represent a diverse river style.

References

Abbado, D., Slingerland, R. and Smith, N.D. (2003) The origin of anastomosis on the Columbia River, British Columbia, Canada, in M.D. Blum, S.B. Marriott and S.M. Leclair (eds) *Fluvial Sedimentology VII, Proceedings of the 7th International Conference on Fluvial Sedimentology*, Special Publication of the International Association of Sedimentologists, 35.

Church, M. (1983) Anastomosed fluvial deposits: modern examples from Western Canada, in J. Collinson and J. Lewin (eds) *Modern and Ancient Fluvial Systems*, 155–168, Special Publication of the International Association of Sedimentologists, 6, Oxford: Blackwell.

Gibling, M.R., Nanson, G.C. and Maroulis, J.C. (1998) Anastomosing river sedimentation in the Channel Country of central Australia, *Sedimentology* 45, 595–619.

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.

Knighton, A.D. and Nanson, G.C. (1993) Anastomosis and the continuum of channel pattern, *Earth Surface Processes and Landforms* 18, 613–625.

Makaske, B. (2001) Anastomosing rivers: a review of their classification, origin and sedimentary products, *Earth-Science Reviews* 53, 149–196.

Nanson, G.C. and Huang, H.Q. (1999) Anabranching rivers: divided efficiency leading to fluvial diversity, in A.J. Miller and A. Gupta (eds) *Varieties of Fluvial Form*, 219–248, Chichester: Wiley.

Nanson, G.C. and Knighton, A.D. (1996) Anabranching rivers: their cause, character and classification, *Earth Surface Processes and Landforms* 21, 217–239.

Smith, D.G. (1973) Aggradation of the Alexandria–North Saskatchewan River, Banff Park, Alberta, in M. Morisawa (ed.) *Fluvial Geomorphology*, 201–219, Binghamton, NY: Publications in Geomorphology, New York State University.

——(1976) Effect of vegetation on lateral migration of anastomosed channels of a glacial meltwater river, *Geological Society of America Bulletin* 86, 857–860.

Smith, D.G. and Smith, N.D. (1980) Sedimentation in anastomosed river systems: examples from alluvial valleys near Banff, Alberta, *Journal of Sedimentary Petrology* 50, 157–164.

Tooth, S.J. and Nanson, G.C. (1999) Anabranching rivers on the Northern Plains of arid central Australia, *Geomorphology* 29, 211–233.

——(2000) The role of vegetation in the formation of anabranching channels in an ephemeral river, Northern plain, arid central Australia, *Hydrological Processes* 14, 3,099–3,117.

Van Niekerk, A.W., Heritage, G.L., Broadhurst, L.J. and Moon, B.P. (1999) Bedrock anastomosing channel systems: morphology and dynamics in the Sabie River, Mpumalanga Province, South Africa, in A.J. Miller and A. Gupta (eds) *Varieties of Fluvial Form*, 33–51, Chichester: Wiley.

Wende, R. and Nanson, G.C. (1999) Anabranching rivers: ridge-form alluvial channels in tropical northern Australia, *Geomorphology* 22, 205–224.

SEE ALSO: avulsion; bedrock channel; braided river; floodplain; meandering

GERALD C. NANSON AND MARTIN GIBLING

ANTHROPOGEOMORPHOLOGY

Anthropogeomorphology is the study of the human role in creating landforms and modifying the operation of geomorphological processes such as weathering, erosion, transport and deposition (see, for example, Brown 1970; Nir 1983;

Goudie 1993). Some landforms are produced by direct anthropogenic processes. These tend to be relatively obvious in form and are frequently created deliberately and knowingly. They include landforms produced by: construction (e.g. spoil tips, bunds, embankments), excavation (e.g. road cuttings, and open-cast and strip mines, etc.), hydrological interference (e.g. reservoirs, ditches, channelized river reaches and canals) and farming (e.g. terraces; see Plate 3).

Landforms produced by indirect anthropogenic processes are often less easy to recognize, not least because they tend to involve not the operation of new processes, but the acceleration of natural processes. They are the result of environmental changes brought about inadvertently by human actions. By removing or modifying land cover – through cutting, bulldozing, burning and grazing – humans have accelerated rates of erosion and sedimentation (see SOIL EROSION). Sometimes the results will be spectacular, for example when major gully systems rapidly develop (see ARROYO, DONGA). By other indirect means humans may create subsidence features (Johnson 1991), cause lake desiccation (Gill 1996), trigger mass movements like landslides, and influence the operation of phenomena like earthquakes through the impoundment of large reservoirs (Meade 1991). Rates of rock weathering may be modified because of the acidification of precipitation caused by accelerated sulphate emissions (see SULPHATION) or because of accelerated salinization in areas of irrigation (Goudie and Viles 1998).

Plate 3 Strip lynchets in Dorset, southern England, are a manifestation of the impact that agricultural activities can have on the geomorphology of slopes. Many of the lynchets are the result of ploughing in medieval times

There are situations where, through a lack of understanding of the operation of geomorphological systems, humans may deliberately and directly alter landforms and processes and thereby set in train a series of events which were not anticipated or desired. There are, for example, many records of attempts to reduce coast erosion by important and expensive hard engineering solutions, which, far from solving erosion problems, only exacerbated them (Bird 1979).

As so often with environmental change, it is seldom easy to disentangle changes that are anthropogenic from those that are natural (Brookfield 1999). There has, for example, been a long-continued debate about the origin of deeply incised gullies, called ARROYOs, which developed in the southwestern United States over a relatively short period in the late nineteenth century. Some workers have championed human actions (e.g. overgrazing) as the cause of this erosion spasm, while others have championed the importance of natural environmental changes, noting that arroyo cutting had occurred repeatedly before the arrival of Europeans in the area. Among the natural changes that could promote the phenomenon are a trend towards aridity (which depletes the cover of protective vegetation) or increased frequencies of high-intensity storms (which generates erosive runoff).

Another example of the complexity of causation is posed by a consideration of the potential causes of loss of land to the sea in coastal Louisiana (Walker *et al.* 1987), something that appears to be proceeding at a rapid rate at the present time. Among the factors that need to be considered are the natural ones of sea-level change, subsidence, progressive compaction of sediments, changes in the locations of deltaic depocentres, hurricane attack and degradation by marsh fauna. Equally, however, one has to consider a range of human actions, including the role that dams and levees have played in reducing the amount and texture of sediment reaching the coast, the role of canal and highway construction and subsidence caused by fluid withdrawals.

In many cases, however, as with the USA Dust Bowl in the 1930s, it is a conjunction in time of human actions (the busting of the sod) with a climatic perturbation (a great drought) that produces change.

The possibility that the build-up of greenhouse gases in the atmosphere may cause enhanced global warming in coming decades has many implications for anthropogeomorphology (see GLOBAL GEOMORPHOLOGY). Increased sea-surface temperatures may change the geographical spread, frequency and wind speeds of hurricanes – highly important geomorphological agents, particularly in terms of river channels and mass movements. Warmer temperatures will cause sea ice to melt and may lead to the retreat of alpine glaciers and the melting of permafrost. Vegetation belts will change latitudinally and altitudinally and this will also influence the operation of geomorphological processes. Changes in temperature, precipitation amounts, and the timing and form of precipitation (e.g. whether it is rain or snow) will have a whole suite of important hydrological consequences. Some parts of the world may become moister (e.g. high latitudes and some parts of the tropics) while other parts (e.g. some of the world's drylands) may become drier. The latter would suffer from declines in river flow, lake desiccation, reactivation of sand dunes and increasing dust storm frequencies.

However, among the most important potential future anthropogeomorphological changes are those associated with sea-level change caused by the steric effect and by the melting of land ice. Low-lying coastal areas (e.g. saltmarshes, mangrove swamps, sabkhas, deltas, atolls) would tend to be particularly susceptible. Moreover, rising sea levels could promote beach erosion, as is suggested by the BRUUN RULE.

Some landscapes – 'geomorphological hot spots' (Goudie 1996) – will be especially sensitive because they are located in areas where it is forecast that climate will change to an above average degree. This is the case, for instance, in the high latitudes of Canada or Russia, where the degree of warming may be 3–4 times greater than the global average. It may also be the case with respect to some areas where particularly substantial changes in precipitation may occur. For example, various scenarios portray the High Plains of the USA as becoming markedly more arid. Other landscapes will be especially sensitive because certain landscape-forming processes are closely controlled by climatic conditions. If such landscapes are close to particular climatic thresholds then quite modest amounts of climatic change can switch them from one state to another.

References

Bird, E.C.F. (1979) Coastal processes, in K.J. Gregory and D.G. Walling (eds) *Man and Environmental Processes*, 82–101, Folkestone: Dawson.

Brookfield, H. (1999) Environmental damage: distinguishing human from geophysical causes, *Environmental Hazards* 1, 3–11.

Brown, E.H. (1970) Man shapes the Earth, *Geographical Journal* 136, 74–85.

Gill, T.E. (1996) Eolian sediments generated by anthropogenic disturbance of playas: human impacts on the geomorphic system and geomorphic impacts on the human system, *Geomorphology* 17, 207–228.

Goudie, A.S. (1993) Human influence in geomorphology, *Geomorphology* 7, 37–59.

——(1996) Geomorphological 'hotspots' and global warming, *Interdisciplinary Science Reviews* 21, 253–259.

Goudie, A.S. and Viles, H.A. (1998) *Salt Weathering Hazards*, Chichester: Wiley.

Johnson, A.I. (ed.) (1991) Land subsidence, *Publication, International Association of Hydrological Sciences*, No. 200.

Meade, R.B. (1991) Reservoirs and earthquakes, *Engineering Geology* 30, 245–262.

Nir, D. (1983) *Man, A Geomorphological Agent, An Introduction to Anthropic Geomorphology*, Jerusalem: Keter.

Walker, H.J., Coleman, J.M., Roberts, H.H. and Tye, R.S. (1987) Wetland loss in Louisiana, *Geografiska Annaler* 69A, 189–200.

A.S. GOUDIE

ANTIDUNE

A symmetrical fluvial BEDFORM produced by near-critical flows, forming in broad shallow channels, and comparable to a sand dune. However, antidunes are more temporary and are less common than dunes. Antidune formation requires a Froude Number (quantifying the relationship between the bedform and flow regime) greater than 0.8, with development often dependent upon channel depth and bed material. Antidunes migrate upstream as sediment is lost from their downstream side more rapidly than it is deposited, though they can also move downstream or remain stationary. Antidunes form directly in phase with standing waves on the water's surface, and are characterized by shallow foresets which dip upstream at an angle of about 10°. They show low resistance to flow and are rare in the rock record, probably due to reworking. Where antidunes are observed in ancient sediments they are characterized by fine, poorly developed laminae.

Further reading

Barwis, J.H. and Hayes, M.O. (1983) Genesis and preservation of antidune stratification in modern and ancient washover deposits, *Association of American Petroleum Geologists' Bulletin* 7(3), 419–420.

Mehrotra, S.C. (1983) Antidune movement, *Journal of Hydraulic Engineering – ASCE* 109, 302–304.

STEVE WARD

APPLIED GEOMORPHOLOGY

The application of geomorphology to the solution of miscellaneous problems, especially to the development of resources and the diminution of hazards (Goudie 2001), to planning, conservation and specific engineering or environmental issues (Brunsden 2002). It incorporates what is sometimes called 'ENGINEERING GEOMORPHOLOGY' (Coates 1971).

In the last three decades applied geomorphology has become a much more central and accepted part of the discipline and a variety of texts have now appeared that review its nature (e.g. Hails 1977; Cooke and Doornkamp 1974; Thorne *et al.* 1997).

The reasons for what Jones (1980: 49) calls this 'significant transformation' are various. He cites three main reasons:

1 An increasing awareness on the part of environmental decision-makers as to the complexity of environmental conditions and the significance of geomorphological hazards (e.g. landslides and floods).
2 Demand from engineers for more information on ground conditions for construction purposes and for engineering geomorphological maps.
3 A decreasing level of insularity among geomorphologists and their feeling that they needed to justify their existence in a society that increasingly measured value in terms of practical achievement.

Other reasons have been the change of emphasis in geomorphology towards the study of contemporary processes and changes at the Earth's surface. The second has been the development of more precise techniques for mapping, monitoring and analysis. A third has been growing awareness of the finite limits to some resources and the importance of a seemingly growing number of environmental crises and catastrophes. Growing urban centres face many serious geomorphological hazards. A fourth stimulus has been the development in various countries of the need for environmental impact assessments.

A major new stimulus to applied geomorphological research has been a concern with global environmental change and the potential consequences of global warming. Matters such as the stability of the Antarctic ice sheets, the susceptibility of permafrost to thermokarst development, the sensitivity of coastal wetlands to sea-level rise, and the possible reactivation of sand seas are some of the major issues that have been investigated and for which land management solutions may be required. At a more local scale, human activities are modifying the rate and extent of particular geomorphological processes including soil erosion, salt weathering and river channel form.

Geomorphologists have a variety of applicable skills which, while they may not individually be unique, as a package are distinctive.

1 'An eye for country' and the ability to interpret landscapes and identify landforms.
2 The ability to interpret and produce maps, for these uniquely effective means of imparting spatial information are central to applied geomorphology. GEOMORPHOLOGICAL MAPPING, based on field surveys and the use of topographic base maps and remote sensing techniques, have for long been used by applied geomorphologists. Cartographic skills have been revolutionized in recent years through the use of new technologies, including differential GPS, GIS, DIGITAL ELEVATION MODELS and LIDAR. Maps are especially important for land use planning and zoning.
3 Competence in the use of techniques to measure the operation of geomorphological processes.
4 Appreciation of the relationships between environmental phenomena. This enables applied geomorphologists to see a site in its broader context and to appreciate that change in one place will have ramifications elsewhere. Thus an engineering scheme (e.g. the construction of an erosion control device on a coastline) can have a range of unintended impacts on slope stability or on downdrift beach nourishment.
5 Recognition of the importance of spatial scale. Geomorphologists appreciate, for example, that rates of sediment yield vary according to the area studied and that small erosion plots may give different orders of magnitude of rates than whole catchment studies in a large basin.
6 Recognition that all places are different and that a practice which may be appropriate in one place may not be appropriate in another. Thus some areas may be peculiarly aggressive while others may be especially sensitive to change. In a permafrost area, for example, there may be profound local differences in permafrost stability because of local soil or microclimatic conditions.
7 Recognition that the landscape is subject to change at all temporal scales and that not only is the present always changing but also that the present is a poor guide to either past or future conditions. An example of such a skill is the recognition of the need to reconstruct long-term discharge records for rivers using a range of dating and sedimentological techniques.
8 Recognition of the importance of human activities and human attitudes. A natural science/social science mix is a unique attribute that will be of particular importance in the field of environmental management (Jones 1980: 70).

The roles of the applied geomorphologist

The various roles of the applied geomorphologist are shown in Table 1.

1 A very basic, but highly important role is to map geomorphological phenomena as a basis for TERRAIN EVALUATION. Landforms, especially depositional ones, may be important sources of useful materials for construction, while maps of slope categories may help in the planning of land use and maps of hazardous ground may facilitate the optimal location of engineering structures.

Plate 4 The main railway line from Swakopmund to Walvis Bay in Namibia illustrates the hazard posed by sand and dune movement. One role of the applied geomorphologist is to identify optimum route corridors and to advise on their management

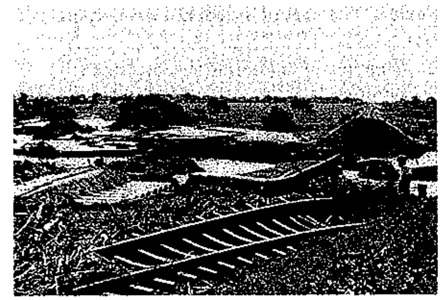

Plate 5 The demolition of a railway line in Swaziland, southern Africa, caused by floods associated with a large tropical cyclone. One role of the applied geomorphologist is to undertake post-event surveys in order to ascertain past discharges

2 Use landforms as the basis for mapping other aspects of the environment, the distribution of which is related to their position on different landforms. This is important because landforms are relatively easily recognized on air photographs and other types of remote sensing imagery. An important example of the use of landforms as surrogates for other phenomena is the use of landform mapping to provide the basis of a soil map through the use of the CATENA concept and soil toposequences.
3 Recognize and measure the speed at which geomorphological change is taking place. Such changes (Table 2) may be hazardous to humans (e.g. coastal retreat, movement of river bluffs, surges of glaciers). By using sequential maps and remote sensing images, archival information or by monitoring processes with appropriate instrumentation, areas at potential risk can be identified, and predictions can be made as to the amount and direction of change.
4 Assess the causes of observed and measured changes and hazards, for without a knowledge of cause, attempts at amelioration and management may have limited success. There is an increasing need to assess the role that humans are playing in modifying rates

Table 1 The roles of the applied geomorphologist

1 Mapping of landforms, resources and hazards
2 Use of maps of landforms as surrogates for other phenomena (e.g. soils)
3 Establishment of rates of geomorphological change by direct monitoring, use of sequential maps, archives, etc.
4 Establishing causes of change
5 Assessment of management options
6 Post-construction assessment of engineering schemes
7 Post-event evaluations (e.g. palaeodischarges)
8 Prediction of future events and changes

Table 2 Examples of geomorphological hazards

Arid zones	*Coastal*
Dune encroachment	Sea-level change
Soil deflation	Dune blowouts
Arroyo formation	Cliff retreat
Dust storms	Saltmarsh siltation
Fan entrenchment	Coastal progradation
Flash floods	Spit growth and breaching
Salt weathering	
Ground subsidence	
Tundra areas	*General*
Thermokarst formation	Mass movement
Frost leave	Karstic collapse
Thaw floods and ice jams	River floods
Glacier surges and glacier dams	Shifting river courses
Avalanches	Lake sedimentation
Jökulhaups	Soil erosion
	Riverbank erosion
	Neotectonic activity

of geomorphological processes, particularly as a result of land-cover changes.

5 Having decided on the speed, location and causes of change, appropriate management solutions need to be adopted. Although the management solution to a particular geomorphological problem may involve the building of an engineering structure (e.g. a sand fence, a sea wall, a check dam, a shelterbelt), these structures may themselves create problems and their relative effectiveness needs to be assessed. The applied geomorphologist may make certain recommendations as to the likely consequences of building, for example, GROYNES to reduce coastal erosion. Examples of engineering solutions having unforeseen environmental consequences, sometimes to the extent that the original problem is heightened and intensified rather than reduced, are all too common, especially in coastal situations (Viles and Spencer 1995). Management issues involve a consideration of ecological issues, such as when one decides on the most appropriate form for a river channelization scheme, and are likely to become increasingly important as decisions have to be made about how to manage the landscape in the face of global climate change. More and more alternatives are being sought to ecologically injurious 'hard engineering' solutions.

6 Related to environmental management and the use of engineering solutions, is the field of assessment of the success of particular schemes. An audit of performance is required as the basis for formulating best practice.

7 Undertake 'after-the-event' surveys. It is important to put on record the magnitude and consequences of extreme events as a basis for improving engineering designs and land-zoning policies. For instance, establishing the Holocene flood histories of rivers by surveying and dating slack water deposits give an important tool for predicting possible future flood peaks, especially in ungauged catchments.

8 This brings us to the final role of the applied geomorphologist, which is to look forward and to predict. When is a particular glacier likely to surge, how long will it take for an irrigation canal to be blocked by a wandering barchan, when is this slope likely to fail, how quickly will this reservoir be rendered useless through sedimentation, will the surface of a delta be built up by fluvial sediment inputs more quickly than sea-level rises? These are examples of where geomorphologists can help to answer questions about the future. Their answers can be based on studies of the past rate of operation of geomorphological processes or by developing their modelling capability.

References

Brunsden, D. (2002) Geomorphological roulette for engineers and planners: some insights into an old game, *Quarterly Journal of Engineering Geology and Hydrogeology* 35, 101–142.
Coates, D.R. (ed.) (1971) *Environmental Geomorphology*, Binghamton: State University of New York.
Cooke, R.U. and Doornkamp, J.C. (1974) *Geomorphology in Environmental Management*, Oxford: Oxford University Press.
Goudie, A.S. (2001) Applied geomorphology: an introduction, *Zeitschrift für Geomorphologie Supplementband* 124, 101–110.
Hails, J.R. (ed.) (1977) *Applied Geomorphology; A Perspective of the Contribution of Geomorphology to Interdisciplinary Studies and Environmental Management*, Amsterdam: Elsevier Science Publishers.
Jones, D.K.C. (1980) British applied geomorphology: an appraisal, *Zeitschrift für Geomorphologie Supplementband* 36, 48–73.
Thorne, C.R., Hey, R.D. and Newson, M. (eds) (1997) *Applied Fluvial Geomorphology for River Engineering and Management*, Chichester: Wiley.
Viles, H.A. and Spencer, T. (1995) *Coastal Problems*, London: Arnold.

A.S. GOUDIE

ARCH, NATURAL

Natural arches are formed when weathering, together with mass collapse, and in arid areas with wind erosion, creates a tunnel through a slab of rock. They can thus be distinguished from NATURAL BRIDGES which are formed by fluvial or marine erosion. They most commonly occur in sandstone, which has sufficient permeability to provide the seepage that promotes weathering, yet which has the necessary cohesion for an arch to develop. Arches are most numerous where long and closely spaced joints have been eroded to form narrow fins of rock that are readily pierced by weathering. These characteristics, and thus a very high concentration of arches, occur in the Entrada and Cedar Mesa Sandstones of the Colorado Plateau.

In strongly bedded rock, widening of the initial tunnel may result in the development of a long slab or lintel. The load of the undercut rock creates tensional stress on the lower face of the slab. If the space continues to grow, the stress may exceed the tensile strength of the rock, causing the slab to collapse. An upward curving form, rather than a slab, will develop where there are curved joints in the rock, or, more commonly, where concave stress patterns in the undercut rock result in minor failures or surface spalling. The curved form of the true arch is much more stable than a slab, because the load is transmitted to the abutments, and virtually the entire structure is in compression. This is so even when the arch is split by joints, for compressive stress on each of the joint-bounded blocks keeps them in place.

Natural arches may take various forms, but will remain stable provided the load is transmitted into the abutments. This condition is met so long as the thrust line of the load remains within the arch. Arches are therefore very stable features. However, continued erosion may result in an unstable form, and the arch may then collapse by folding in on itself at several hinge points. Erosional weakening of the abutments into which the load is transmitted can also cause an arch to collapse. Conversely, rock pinnacles transmitting a vertical load down into the abutments, or natural buttresses supporting them laterally, increase the stability of the arch.

Especially where joints are widely spaced, the hollow developed in a cliff may not penetrate through the rock mass. Instead of a true arch, an alcove or apse develops. These are much more widespread than arches, but they form in essentially the same manner, being particularly well developed where seepage issues from massive sandstone.

Further reading

Robinson, E.R. (1970) Mechanical disintegration of the Navajo sandstone in Zion Canyon, Utah, *Geological Society of America Bulletin* 81, 2,799–2,806.
Young, R. and Young, A. (1992) *Sandstone Landforms*, Berlin: Springer.

R.W. YOUNG

ARÊTE

A landform composed of a fretted, steep-sided ridge that separates valley or cirque GLACIERS. Arêtes are the result of glacial undercutting – basal sapping – of rock slopes. They are common whenever mountains and peaks rise above glaciers, as in the case with NUNATAKS.

A.S. GOUDIE

ARMOURED MUD BALL

Roughly spherical lumps of cohesive sediment which generally have a diameter of a few centimetres (Bell 1940). They are also called mud balls, mud pebbles, pudding balls, till balls and clay balls. Many examples are lumps of clay or cohesive mud that have been eroded by vigorous currents from stream beds or banks. They often occur in areas of badland topography and along ephemeral streams, but can also be found on beaches (Kale and Awasthi 1993), in tidal channels, and as ice-transported debris dumped on the seafloor (Goldschmidt 1994).

References

Bell, H.S. (1940) Armoured mud balls: their origin, properties and role in sedimentation, *Journal of Geology* 48, 1–31.
Goldschmidt, P.M. (1994) Armoured and unarmoured till balls from the Greenland sea-floor, *Marine Geology* 121, 121–128.
Kale, V.S. and Awasthi, A. (1993) Morphology and formation of armoured mud balls on Revadanda Beach, Western India, *Journal of Sedimentary Petrology* 63, 809–813.

A.S. GOUDIE

ARMOURING

'Armouring', the process whereby a clastic deposit develops a surface layer that is coarser than the substrate, is most commonly associated with warm deserts and gravel-bed rivers (see BOULDER PAVEMENT, STONE PAVEMENT, FLUVIAL ARMOUR). DEFLATION, the removal of fine-grained material by wind; the winnowing of fines by surface wash (see SHEET EROSION, SHEET FLOW, SHEET WASH); the upward migration of coarse particles, as a result of alternate wetting and drying and the associated swelling and shrinking of fine debris, or FREEZE–THAW CYCLE activity at high altitudes; the upward displacement of gravel clasts as fines accumulate; and the preferential weathering and breakdown of coarse debris at depth have all been proposed as mechanisms that produce concentrations of coarse particles at the ground surface in desert environments (Cooke 1970; Dan *et al.* 1982; McFadden *et al.* 1987). The process(es) operating in any given location depend on climate, geomorphic setting, and the nature of the clastic particles and local soils. The gravel clasts involved may be produced by mechanical weathering of the local bedrock, or be of fluvial origin. In rivers, armouring may involve the concentration of coarse clasts at the base of the active layer or the preferential winnowing of finer sediment from the surface during degradation; and vertical winnowing during active BEDLOAD transport which compensates for the disparity in mobility between coarse and fine particles (Andrews and Parker 1987; Parker and Sutherland 1990). Size segregation which produces concentrations of large particles at the surface, also occurs in gravity-driven, granular mass flows, including DEBRIS FLOWS and PYROCLASTIC FLOW DEPOSITS. Segregation mechanisms include size percolation, in which fine grains infiltrate beneath coarser particles; size exclusion, where coarse grains are excluded from narrow, convective downwellings; and cascading segregation, in which larger particles roll more rapidly downslope than small ones (Shinbrot and Muzzio 2000; Vallance and Savage 2000).

References

Andrews, E.D. and Parker, G. (1987) Formation of a coarse surface layer as the response to gravel mobility, in C.R. Thorne, J.C. Bathurst and R.D. Hey (eds) *Sediment Transport in Gravel-Bed Rivers*, 269–300, Chichester: Wiley.
Cooke, R.U. (1970) Stone pavements in deserts, *Annals of the Association of American Geographers* 60, 560–577.
Dan, J., Yaalon, D.H., Moshe, R. and Nissim, S. (1982) Evolution of reg soils in southern Israel and Sinai, *Geoderma* 28, 173–202.
McFadden, L.D., Wells, S.G. and Jercinovich, M.J. (1987) Influences of eolian and pedogenic processes on the origin and evolution of desert pavements, *Geology* 15, 504–508.
Parker, G. and Sutherland, A.J. (1990) Fluvial armor, *Journal of Hydraulics Research* 28, 529–544.
Shinbrot, T. and Muzzio, F.J. (2000) Nonequilibrium patterns in granular mixing and segregation, *Physics Today* 53, 25–30.
Vallance, J.W. and Savage, S.B. (2000) Particle segregation in granular flows down chutes, in A. Rosato and D. Blackmore (eds) *International Union of Theoretical and Applied Mechanics Symposium on Segregation in Granular Flows*, Dordrecht, The Netherlands, 31–51.

BASIL GOMEZ

ARROYO

An incised valley bottom, particularly in the western USA (Figure 7), where many broad valleys and plains became deeply incised with valley-bottom gullies (arroyos) over a short period between 1865 and 1915, with the 1880s being especially important (Cooke and Reeves 1976). The arroyos can be cut as deeply as 20 m, be over 50 m wide and tens or even hundreds of kilometres long. There has been a long history of debate as to the causes of incision (Elliott *et al.* 1999) and an increasing

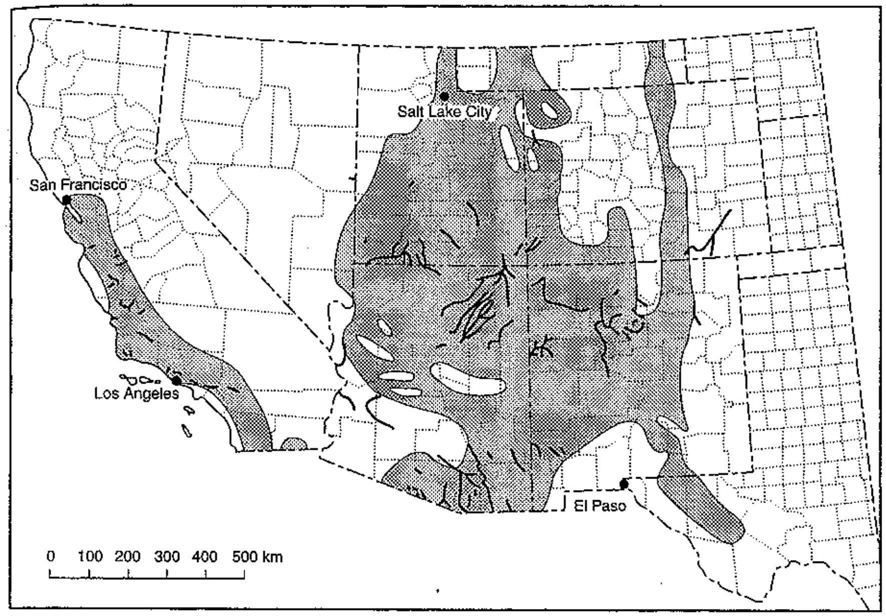

Figure 7 The distribution of arroyos in the southwestern USA (shaded), showing the course of some large examples (dark lines)

appreciation of the scale and frequency of climatic changes in the Holocene (McFadden and McAuliffe 1997), which could have led to changes in channel and slope behaviour. For example, Waters and Haynes (2001) have argued that arroyos first appeared in the American south-west after *c.*8,000 years ago, and that a dramatic increase in cutting and filling episodes occurred after *c.*4,000 years ago. They believe that this intensification could be related to a change in the frequency and strength of El Niño events.

Many students of this phenomenon have believed that human actions caused the entrenchment, and the apparent coincidence of white settlement and arroyo development in the late nineteenth century tended to give credence to this viewpoint. The range of actions that could have been culpable is large: timber-felling, overgrazing, cutting grass for hay in valley bottoms, compaction along well-travelled routes, channelling of runoff from trails and railways, disruption of valley-bottom sods by animals' feet, and the invasion of grasslands by scrub.

On the other hand, study of the long-term history of the valley fills shows that there have been repeated phases of aggradation and incision and that some of these took place before the influence of humans could have been a significant factor. Elliott *et al.* (1999) recognize various Holocene phases of channel incision at 700–1,200 BP, 1,700–2,300 BP and 6,500–7,400 BP.

A climatic interpretation was advanced by Leopold (1951), which involved a change in rainfall intensity. He indicated that a reduced frequency of low-intensity rains would weaken the vegetation cover, while an increased frequency of heavy rains at the same time would increase the incidence of erosion. Balling and Wells (1990), working in New Mexico, attributed early twentieth-century arroyo trenching to a run of years with intense and erosive rainfall that succeeded a phase of drought conditions in which the protective ability of the vegetation had declined. Large floods have also been important causative agents (Hereford 1986). Erosion and entrenchment result from a larger flood regime, with streams having a large sediment transport capacity. With lower flood regimes, a reduction in channel width and sediment storage occur, but if there are no floods, no alluviation of floodplains is possible. It is also possible, as Schumm *et al.* (1984) have pointed out, that arroyo incision could result from neither climatic change nor human influence. It could be the result of some intrinsic natural geomorphological threshold (see THRESHOLD, GEOMORPHIC) (such as stream gradient) being crossed. Under this argument, conditions of valley-floor stability decrease slowly over time until some triggering event initiates incision of the previously 'stable' reach.

It is possible that arroyo incision and alluviation result from a whole range of causes (Gonzalez 2001), that the timing of events will have varied from area to area and that individual arroyos will have had unique histories.

References

Balling, R.C. and Wells, S.G. (1990) Historical rainfall patterns and rainfall activity within the Zuni River drainage basin, New Mexico, *Annals of the Association of American Geographers* 80, 603–617.

Cooke, R.U. and Reeves, R.W. (1976) *Arroyos and Environmental Change in the American South-west*, Oxford: Clarendon Press.

Elliott, J.G., Gillis, A.C. and Aby, S.B. (1999) Evolution of arroyos: incised channels of the southwestern United States, in S.E. Darby and A. Simon (eds) *Incised River Channels*, 153–185, Chichester: Wiley.

Gonzalez, M.A. (2001) Recent formation of arroyos in the Little Missouri Badlands of southwestern Dakota, *Geomorphology* 38, 63–84.

Hereford, R. (1986) Modern alluvial history of the Paria River drainage basin, southern Utah, *Quaternary Research* 25, 293–311.

Leopold, L.B. (1951) Rainfall frequency: an aspect of climate variation, *Transactions of the American Geophysical Union* 32, 347–357.

McFadden, L.D. and McAuliffe, J.R. (1997) Lithologically influenced geomorphic responses to Holocene climatic changes in the southern Colorado Plateau, Arizona: a soil-geomorphic and ecologic perspective, *Geomorphology* 19, 303–332.

Schumm, S.A., Harvey, M.D. and Watson, C.C. (1984) *Incised Channels: Morphology, Dynamics and Control*, Littleton, CO: Water Resources Publications.

Waters, M.R. and Haynes, C.V. (2001) Late Quaternary arroyo formation and climate change in the American southwest, *Geology* 29, 399–402.

A.S. GOUDIE

ASPECT AND GEOMORPHOLOGY

As the sun moves across the sky, through the course of each day and through the seasons, the intensity of short wave radiation at a point on the hillside changes. At night, there is little radiation. In the daytime, radiation is greatest when the sun is un-obscured, and not reduced by cloud cover or where the hillside is shaded by surrounding hills. Because, north of the equator, the sun is highest in the sky towards the south, sunny south-facing slopes receive more short wave radiation than north-facing slopes, while east- and west-facing slopes receive intermediate amounts, east-facing slopes receiving more in the mornings and west-facing in the evenings. In the southern hemisphere, relationships are exchanged, north-facing slopes receiving most radiation, although east-facing slopes still face the morning sun.

Some solar radiation is lost in passing through the atmosphere, partly through the scattering which gives blue sky light, and much more if there are clouds in front of the sun. The radiation from both a cloudy sky and a clear blue sky is diffuse, and comes from all directions, although some light is lost by shading in deep valleys. The direct beam of the un-obscured sun is strongly directional, and its intensity on the surface is directly proportional to the cosine of the angle between the sun's rays and a perpendicular to the slope surface. Thus solar radiation is highest where rays fall squarely on the surface, and is greatly reduced when the rays graze the surface.

The sun's path through the sky changes in a regular way through the year, so that the amount of radiation on a hillside can be computed trigonometrically, from the latitude, the slope gradient and the direction in which the slope faces.

The sun's azimuth Φ (bearing to the sun's position in the sky) and elevation θ (angle above the horizon) can be calculated with reasonable accuracy as:

$$\sin\theta = \sin\lambda\sin\beta - \cos\lambda\cos\beta\cos\gamma$$

$$\tan\Phi = \frac{\cos\beta\sin\gamma}{\sin\lambda\cos\beta\cos\gamma + \cos\lambda\sin\beta}$$

where λ is the latitude in degrees North, β is the sun's declination ~ $-23.5\cos J$ on Julian day J (0–360) and γ is $15h$ at hour h (0–24 hr local sun time). Even under a clear sun, some light is diffused (about 15 per cent under unpolluted skies) to provide blue sky light. Corrections must also be made for cloudiness and shading by any hills which form the local horizon. Making these calculations, Figure 8 shows that the difference in radiation received from clear skies on north- and south-facing slopes is greatest at about 60° latitude, but because cloudiness also increases with latitude on the continents from 30°, particularly in summer, the actual difference in radiation received is greatest at latitudes of 30°–40°.

Aspect affects geomorphology through the contrasts in radiation, most strongly between north- and south-facing slopes, which leads to differences in hydrology and sediment transport rates. Table 3 summarizes the main differences for the northern hemisphere, and north and south should be consistently exchanged for the southern hemisphere.

The effect of aspect differences is generally to create differences in the intensity of geomorphic processes between the two opposing hillsides. For example the greater radiation on south-facing semi-arid slopes increases evapotranspiration rates, so that water stress occurs in vegetation more quickly after rain. As a result, the vegetation cover is sparser and the species more drought adapted. Sparse vegetation encourages greater crusting of the soil surface, more overland flow runoff and more erosion by wash erosion. On north-facing slopes, soil moisture is maintained after rain for a longer period, so that humid vegetation can grow, usually providing greater ground cover, and better conditions for soil accumulation. Although these conditions improve infiltration rates and reduce overland flow and wash erosion, they can also provide better conditions for mass movements due to the greater depth of soil and higher moisture content.

In the short term, an increase in erosion may lead to steepening of the slope profile, but the longer term implications, as slope profiles approach some form of equilibrium, are less clear, although process differences due to aspect are commonly associated with ASYMMETRIC VALLEYS. Where there is pronounced slope asymmetry, short steep slopes on one side of the valley are matched by longer and gentler slopes on the opposite side. Two factors influence the form of the asymmetric valley cross-section. First, sediment transport depends on both slope length and slope gradient, so that the steeper slope does not necessarily deliver the more sediment. Second, at equilibrium, the valley form may not only be cutting vertically downwards, but may also be migrating laterally. For both these reasons, the hillside with the more intense process activity, due to aspect differences, may not become the gentler slope to compensate for its more intense geomorphological activity. Observations of semi-arid slopes generally suggest that radiation differences tend to maintain steep bedrock slopes on south-facing aspects, and gentler slopes mantled with soil and vegetation on north-facing aspects,

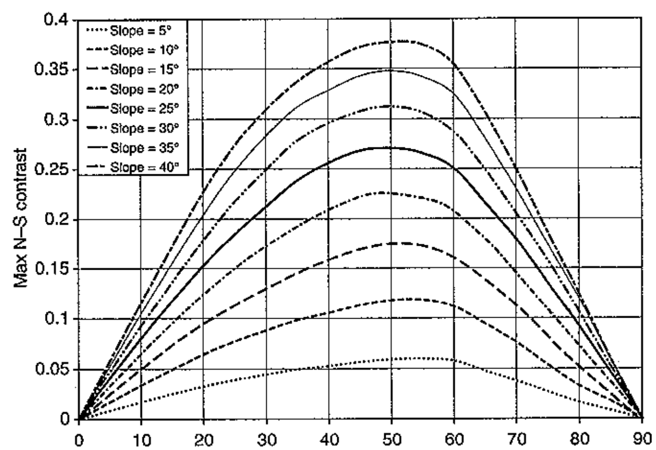

Figure 8 Difference in total annual radiation between north and south-facing slopes for clear sky conditions

Table 3 Summary of effects of aspect differences

Climatic regime	North-facing	South-facing	Geomorphic impact
Very cold (arctic or high altitude)	Permanently frozen	Some freeze-thaw	Greater solifluction and other activity on S-facing slopes
Moderately cold	Some freeze-thaw	Mainly unfrozen	Greater disturbance of vegetation and solifluction on N-facing slopes
Moist temperate	Cooler and moister	Warmer and dryer	Where water is not limiting, differences due to aspect are weak
Warm semi-arid	Cooler and moister	Warmer and dryer	S-facing slopes have sparser and more xeric vegetation, and greater runoff and erosion

although the strength of asymmetry is affected by a number of other factors, particularly geological structure and the meandering activity of rivers.

Further reading

Kirkby, M.J., Atkinson, K. and Lockwood, J.G. (1990) Aspect, vegetation cover and erosion on semi-arid hillslopes, in J.B. Thornes (ed.) *Vegetation and Erosion*, 25–39, Chichester: Wiley.
Robinson, N. (1966) *Solar Radiation*, Amsterdam: Elsevier.

MIKE KIRKBY

ASTROBLEME

The term *astrobleme* (literally 'star-wound') was introduced by Robert S. Dietz (1960) in reference to ancient erosional scars, usually circular in outline, that form on the Earth's surface through the impact of a cosmic body. This origin was recognized because of the presence of highly disturbed rocks that display evidence of intense shock (Dietz 1961). In the early debates over the origins of these features, it was not clear whether the intense pressures responsible for the disturbed rocks resulted from a bolide (an exploding meteor or comet) or from a volcanic explosion. Structures formed in the latter manner were termed *cryptovolcanic* by Branco and Frass (1901). However, the nongenetic term, *cryptoexplosion structure* (Dietz 1959), is preferred when the origin is uncertain. Nevertheless, modern research methods can nearly always confirm or reject an origin by meteor or comet impact.

The sites of relatively recent impacts on Earth, Quaternary to late Tertiary in age, will generally preserve the morphology of impact craters that is so commonly observed on the surfaces of other rocky planetary bodies in the solar system. The lack of long-term preservation of distinctive crater morphologies on the Earth's surface is the result of long-acting and relatively rapid erosional and depositional processes, when compared to circumstances on the other planetary objects. The ancient eroded impact structures of Earth (Plate 6) include circular features that are much larger than the better preserved, young impact craters. Debates over the cryptovolcanic versus impact origin of these large features raged until about the 1960s, when mineralogical studies confirmed that one of these structures, the Ries Kessel in Germany, was clearly the result of a large impact.

During the impact cratering process, immense pressures are imparted on target rocks by the high-velocity projectile. The highest pressures vaporize and melt rocks upon their release. Indeed, some large astroblemes, like Ries Kessel, are associated with huge amounts of impact melt, which early workers found difficult to distinguish from igneous rocks. Somewhat lower pressures are responsible for the metamorphic alteration of quartz to coesite and stishovite, minerals which do not form in the tectonic and volcanic processes of Earth's interior. Even lower pressures produce distinctive planar features in crystals, shocked quartz, and a distinctive cone-in-cone fracture pattern in target rocks, called *shatter cones*. The study of such features, along with their structural and geological settings, has led to the discovery of well over a hundred terrestrial astroblemes over the last several decades.

Plate 6 Central uplift of the deeply eroded Gosses Bluff impact structure, an astrobleme in central Australia. The bluff comprises a ring of resistant sandstone, about 5 km in diameter, that was uplifted in the centre of a much larger transient crater created during the early Cretaceous (Milton *et al.* 1972). The larger structure has a diameter of about 22 km, but it is has been eroded to a nearly level plane. An ancient, higher planation surface bevels the crests of the sandstone ridges that mark the central uplift

Perhaps the most famous astrobleme is the Chicxulub structure, which is buried beneath cover rocks at the northern end of the Yucatan Peninsula, Mexico (Hildebrand *et al.* 1991). The recognition of this feature and its significance illustrates the highly interdisciplinary character of planetary science studies in application to the Earth. The story begins with the discovery in the late 1970s of an enrichment in the element iridium in a 3-cm thick layer of clay at the Cretaceous–Tertiary boundary in a thick section of marine sediments at Gubbio, Italy (Alvarez *et al.* 1980). This geochemical anomaly led the discoverers to propose that a 10-km diameter comet or asteroid collided with Earth 65 million years ago, ending the Cretaceous era and causing one of the most extensive mass extinctions of organisms in geological history, including the demise of the dinosaurs. This was indeed a provocative hypothesis, of immense potential importance to our understanding of Earth history. How could it be verified?

The iridium anomaly was subsequently identified at numerous other Cretaceous–Tertiary boundary sites around the world. Associated with the iridium were other, somewhat exotic elements in concentrations typical for chrondritic meteorites, as would be expected from the composition of the impactor. Also found were shocked quartz grains, stishovite, coesite and small glass spherules. The latter are interpreted as microtektites. Long considered a geological curiosity, relatively large, pebble-sized tektites have been found over extensive surfaces in local regions. They are clearly melted silicates, but their streamlined shapes showed that they had fallen through the atmosphere. Modern understanding of impact cratering mechanics shows that tektites are droplets of impact melt that achieve widespread ballistic dispersal from very large impact events.

The geochemical evidence all pointed to an object that would have produced a crater about 200 km in diameter, which was considerably larger than any astrobleme that had yet been identified on Earth. By following various indicators of proximity to the impact source, including tsunami deposits, tektite sizes and other features of world Cretaceous–Tertiary boundary deposits, the assembled evidence all pointed to the Caribbean and Gulf of Mexico as the likely target area. Interest then moved toward a previously obscure circular structural anomaly in northern Yucatan. The Chicxulub feature is about 180 km in diameter. Though buried, it has surface expression in a ring of cenotes (karstic sinkholes), and it is well displayed in geophysical surveys of the subsurface structure.

The discovery of astroblemes is accelerating. New features are being found on the ocean floor, aided by the extensive exploration for hydrocarbon resources. The techniques for identifying these anomalous forms make use of classical geomorphological reasoning. Moreover, it is now clear that impact cratering, the most prevalent geomorphological process on the rocky planetary bodies of the solar system, is not so rare on Earth as was once believed. It is just that the immense timescales involved for the larger impacts means that their landform consequences mostly appear as eroded, buried, and/or exhumed features that are intimately associated with the Earth's long-term geological record.

References

Alvarez, L.W., Alvarez, W., Asaro, W. and Michel, H.V. (1980) Extraterrestrial cause for the Cretaceous–Tertiary extinction, *Science* 208, 1,095–1,108.
Branco, W. and Frass, E. (1901) *Das vulcanische ries bei Nördlingen in seiner Bedeutung für Fragen der Allgemeinen Geologie*, Berlin: Akademie der Wissenschaften.

Dietz, R.S. (1959) Shatter cones in crytoexplosion structures (meteorite impact?), *Journal of Geology* 67, 496–505.
——(1960) Meteorite impact suggested by shatter cones in rock, *Science* 131, 1,781–1,784.
——(1961) Astroblemes, *Scientific American* 205, 50–58.
Hildebrand, A.R., Penfield G.T., Kring, D.A., Pilkington, M., Camargo, Z.A., Jacobsen, S.B. and Boynton, W.W. (1991) Chicxulub Crater: a possible Cretaceous/Tertiary boundary impact crater on the Yucatan Peninsula, Mexico, *Geology* 19, 867–871.
Milton, D.J., Barlow, B.C., Brett, R., Brown, A.Y., Glikson, A.Y., Manwaring, E.A. *et al.* (1972) Gosses Bluff impact structure, Australia, *Science* 175, 1,199–1,207.

SEE ALSO: crater; cryptovolcano; extraterrestrial geomorphology

VICTOR R. BAKER

ASYMMETRIC VALLEY

In very few valleys are the profiles of the opposite sides exact mirror images about the axis of the thalweg; the geomorphological definition of valley asymmetry, however, requires substantial differences in the shape and/or steepness of the two hillsides. This asymmetry may be localized, e.g. where a meander creates a river cliff opposite a slip-off slope, or valley-wide, e.g. in the case of the UNICLINAL SHIFTING characteristic of scarp and vale scenery. The ultimate in asymmetry is the case of 'one-sided' valleys, such as those of glaciated regions where the missing side was once provided by an ice sheet.

Asymmetry can, then, be the product of a whole range of circumstances relating to the orientation of valley axes and hillsides with respect to both the underlying geology and the past and present subaerial processes. Kennedy (1976) lists eight factors which have been considered to produce valley asymmetry: Coriolis force; differences in insolation and precipitation receipts; differences in slope dimensions; variable lithology; geologic structure; warping; evolution of the drainage net; and glaciation. Of these the role of geologic structure and of aspect-induced variations (see ASPECT AND GEOMORPHOLOGY) in microclimate are the two most commonly attributed causes of asymmetry.

To deal with geologic structure: faulting is evidently capable of producing dramatic asymmetry, either by opposing a fault scarp to a lower-angle hillside, or by creating hillsides with contrasting lithologies. More generally, it is accepted that the low-angle dips of domes such as the English Weald can lead to preferential down-dip migration of rivers, in the process of uniclinal shifting, resulting in broad and broadly asymmetric valleys. Whilst this is a widely observed geologic control, the question of any more general influence exerted by the dip of beds on the movement of stream channels has never been fully explored. M.J. Selby's ROCK MASS STRENGTH classification includes the dip of joints (and bedding planes), but his concept of strength-equilibrium slopes excludes those undercut by streams (1993: 104).

Far more attention has been directed towards the role of microclimatic variability and the asymmetry of slope processes which results. This was explicitly tested by A.N. Strahler (1950) in his quantitative investigation of the Davisian explanatory trio of 'structure, process and stage'. Working in the Verdugo Hills, California, Strahler found that marked vegetation contrasts between north- and south-facing hillsides were not reflected in significant angular differences. This study was extended and refined by M.A. Melton (1960) who revealed statistically significant asymmetry associated both with profile orientation and with the location of stream channels in the Laramie Mountains, Wyoming; the steepening of undercut profiles was shown to be additively linked to that associated with slope aspect (north-facing steeper).

Kennedy (1976) summarizes evidence for the presence or absence of localized and valley-wide asymmetry in seven areas of North America, ranging from 69°N to 31°N. There is no simple pattern, with the exception of the greater prevalence of valley-wide asymmetry in basins whose axes trend east-west, rather than north-south. This suggests strongly that it is the radiation balance, rather than differential precipitation inputs – at least in these cases – which is crucial to the development of process asymmetry. What is of particular interest, however, is the finding (Kennedy and Melton 1972) that an area of modern permafrost (the Caribou Hills, Northwest Territory) shows distinct, topographically determined cases where either north-facing or south-facing slopes are steeper. This must cast some doubt on the persistent attempts (cf. French 1996) to identify distinctive 'periglacial' asymmetry in terms of the orientation of steeper slopes. Kennedy found steeper north-facing slopes as far south as Kentucky (38°N), where it would seem improbable that they represent any legacy of periglaciation.

If there is any generalization to be made about the role of aspect in inducing valley-wide asymmetry, it is probably that it will develop in cases where the overall moisture balance is in some sense marginal: where this is the case, small topographic differences (or – cf. Schumm 1956 – lithologic ones) may create relatively dramatic variations in infiltration, runoff and mass movements and, ultimately, angular differences. That said, one must largely agree with Selby's assessment: 'few [studies] are based on critical examination of all slope units . . . and even in those that are, it has proved impossible to relate hillslope asymmetry to processes, because hillslopes develop over long periods' (Selby 1993: 289–290).

References

French, H.M. (1996) *The Periglacial Environment*, 2nd edition, Harlow Addison-Wesley.
Kennedy, B.A. (1976) Valley-side slopes and climate, in E. Derbyshire (ed.) *Geomorphology and Climate*, 171–201, Chichester: Wiley.
Kennedy, B.A. and Melton, M.A. (1972) Valley asymmetry and slope forms of a permafrost area in the Northwest Territories, Canada, *Special Publication, Institute of British Geographers* 4, 107–121.
Melton, M.A. (1960) Intravalley variation in slope angles related to microclimate and erosional environment, *Geological Society of America Bulletin* 71, 133–144.
Schumm, S.A. (1956) The role of creep and rainwash on the retreat of badland slopes, *American Journal of Science* 254, 693–706.
Selby, M.J. (1993) *Hillslope Materials and Processes*, 2nd edition, Oxford: Oxford University Press.
Strahler, A.N. (1950) Equilibrium theory of erosional slopes, approached by frequency distribution analysis, *American Journal of Science* 248, 673–696.

Further reading

Bloom, A.L. (1998) *Geomorphology: A Systematic Analysis of Late Cenozoic Landforms*, 3rd edition, Upper Saddle River, NJ: Prentice Hall.
Parsons, A.J. (1988) *Hillslope Form*, London: Routledge.
Tricart, J. (1963) *Géomorphologie des régions froides*, Paris: Presses Universitaires de France.

SEE ALSO: aspect and geomorphology; cross-profile, valley; rock mass strength

BARBARA A. KENNEDY

ATOLL

Atolls are generally sub-circular rings of CORAL REEF surrounding a lagoon with no dry land other than occasional islands (called *motu*) made from sand and gravel-sized detritus thrown up on the reef during storms (Nunn 1994). The word 'atoll' should strictly be applied only to the reef and lagoon (as it is here) but is sometimes used more loosely to refer to *motu*.

On first encounter, it may come as a surprise how ancient many atolls are. In the Pacific, where some of the world's oldest atolls exist, many have reef foundations dating from at least the Oligocene. It may be even more of a surprise to hear how apparently strong such organic structures are, remaining intact despite the continuous buffering of storm waves, earthquakes and even nuclear weapons tests. Yet as we learn more about the structural history of such islands, so it is becoming clear that atolls do, even without such stresses, occasionally experience catastrophic collapses. Thus Johnston Atoll in the central Pacific, where the US chemical weapon stocks are being destroyed, lost its southern flank in a series of huge landslides predating its discovery by humans. On the other hand, part of Moruroa Atoll in French Polynesia, where 98 subterranean tests of nuclear bombs were carried out between 1981 and 1991, has subsided as a direct consequence and growing concern exists about the stability of the remainder – and the possibility of radioactive residues leaking out into the ocean from the test chambers (Keating 1998).

Atoll origins

The classic exposition of atoll origins was that given by Charles Darwin who, having reached Tahiti in 1835 on the *Beagle*, climbed the slopes above Papeete and, looking across at neighbouring Mo'orea Island surrounded by its barrier reef, realized that if the island disappeared, only the REEF would remain. Thus an atoll would be formed. So, even before he had seen an atoll, Darwin set out his Theory of Atoll Development which involved the upward growth of a coral reef in response to the subsidence of its foundations (Darwin 1842). Darwin figured, with considerable prescience, that it was the tendency of ancient volcanic islands in the world's oceans to subside but that their coral fringe could stay alive only if it was able to grow upwards at the same rate. Thus modern atoll reefs (and FRINGING REEFS and CORAL REEFS) had only veneers of living coral growing atop a coral framework composed largely of the skeletal remains of dead hermatypic (reef-building) corals.

For Darwin, atoll lagoons were places where the volcanic foundations of atolls were buried by reef detritus washed over the reef during storms. Organic and mechanical forces combined to make these lagoon sediments finer over time.

What Darwin was unaware of was that sea level had oscillated with amplitudes of 100 m or more during much of the past few million years and that, although this fact did not invalidate his basic model (which is still regarded as essentially correct), sea-level change needs to be incorporated into models of atoll formation. During every period of low sea level, atolls would be converted into high limestone islands analogous to modern Niue Island in the central Pacific and others. The surfaces of these limestone islands would be reduced by KARST erosion and, when sea level rose once again at the end of the low sea-level stand, the reef would begin growing on the reduced surface (Purdy 1974).

It is worth taking a closer look at what would happen when the reef began growing once again on these surfaces during postglacial periods, taking the last as an example. In places where oceanographic and other conditions were most favourable, the upward growth of coral reef was able to 'keep up' with sea-level rise. Yet in most places, it seems that coral could not grow upwards as fast as sea level rose and that only later did the reef surface 'catch up' with sea level. In some cases, reef upgrowth was altogether too slow to keep pace with sea-level rise and the reef 'gave up' resulting in the formation of a drowned atoll (Neumann and MacIntyre 1985). The presence of drowned atolls in many parts of the Pacific and Indian oceans may be a result of Holocene reefs 'giving up' an unequal race as well as their latitude, a proxy for calcium carbonate production (Grigg 1982) and other conditions (Flood 2001).

Atoll forms

It seems most likely that the aerial form of any atoll reflects the subsurface form of the island from which it grew most recently (Purdy 1974). But within that general principle, there is considerable variation of atoll form which cannot be so readily explained.

Like barrier reefs, atoll reefs tend to be broader and more biotically diverse along their windward sides. These are also the places where atoll *motu* are generally more abundant. Thus atolls in the central Pacific easterly wind and swell belt tend to have broader reefs and more (extensive and continuous) *motu* along their eastern sides than their western sides. In contrast, Diego Garcia Atoll in the central Indian Ocean, which experiences a reversal in swell direction every six months, has a symmetrical reef with a continuous *motu*.

On some atolls, particularly those with completely enclosed lagoons, *motu* are extending lagoonwards and beginning to fill in lagoons. Some islands in Tuvalu, such as Nanumaga, which have only a few small lakes and depressions in their central parts are thought to have formed in this manner.

Humans and atolls

It is clear that most atolls only became habitable in the late Holocene because a fall of sea level exposed the tops of atoll reefs which then became foci for the accumulation of sediment dredged up from reef-talus slopes by large waves to form *motu*. Thus the existence of atoll *motu* is clear evidence for the occurrence of a higher-than-present sea level during the middle Holocene, about 4,000 cal. yr BP (Nunn 1994).

Humans have occupied many atolls continuously since that time but stress comes today from many sources. Not only has atoll life become more difficult as populations have increased and demands on naturally resource-poor environments have become more complicated, but today the fabric of atoll islands is threatened by sea-level rise. The rise of sea level during the twentieth century has caused erosion along many atoll shorelines, although often a direct link has been difficult to demonstrate because of a lack of understanding of lagoon sediment dynamics and because of the construction of artificial structures like causeways to link *motu*. In this regard, the effects of creating or enlarging reef passes to enable larger vessels to enter atoll lagoons has created problems for some atoll communities (Nunn 1994).

For the future, there is a widespread perception that the projected rise in sea level in the twenty-first century will result both in the comprehensive destruction of many atoll islands and in many atoll dwellers becoming environmental refugees in consequence. There is certainly cause for concern; one of the best geomorphological studies of recent years (Dickinson 1999) showed that the sea was currently attacking the lithified foundation of Funafuti Atoll in Tuvalu but that soon it was likely that this level would be overtopped and the sea would find itself eroding only the unconsolidated cover of the *motu* resulting in their rapid removal.

References

Darwin, C.R. (1842) *Structure and Distribution of Coral Reefs*, London: Smith, Elder.
Dickinson, W.R. (1999) Holocene sea-level record on Funafuti and potential impact of global warming on central Pacific atolls, *Quaternary Research* 51, 124–132.
Flood, P.G. (2001) The 'Darwin Point' of Pacific Ocean atolls and guyots: a reappraisal, *Palaeogeography, Palaeoclimatology, Palaeoecology* 175, 147–152.
Grigg, R.W. (1982) Darwin Point: a threshold for atoll formation, *Coral Reefs* 1, 29–34.
Keating, B.H. (1998) Nuclear testing in the Pacific from a geologic perspective, in J.P. Terry (ed.) *Climate and Environmental Change in the Pacific*, 113–144, Suva: SSED, The University of the South Pacific.
Neumann, A.C. and MacIntyre, I. (1985) Reef response to sea-level rise: keep-up, catch-up or give-up, in *Proceedings of the 5th International Coral Reef Congress* 3, 105–110.
Nunn, P.D. (1994) *Oceanic Islands*, Oxford: Blackwell.
Purdy, E.G. (1974) Reef configuration: cause and effect, *Society of Economic Paleontologists and Mineralogists, Special Publication* 18, 9–76.

PATRICK D. NUNN

AVALANCHE BOULDER TONGUE

Avalanche boulder tongues are distinctive accumulations of coarse debris resulting from the long continued, downslope transport of debris by snow avalanches (see AVALANCHE, SNOW). Two basic forms were identified (Rapp 1959). Fan tongues are thin veneers of angular debris in the avalanche runout zone. Many larger boulders and vegetation have a scattered surface cover of smaller 'perched' boulders that have been let down in precarious positions from an ablating snow cover. These fans may extend for several hundred metres across slopes of as little as 8°. Similar low-angled fans may also result from the activity of SLUSHFLOW in subarctic environments.

Where avalanches run across accumulations of loose debris (e.g. SCREE slopes below major couloirs) they erode loose debris from the upper surface of these slopes and, redistributing it downslope, produce a raised tongue of debris extending from the base of the original deposit. These 'roadbank' tongues have a pronounced basal concavity and are often asymmetric in cross section with a smoothed bevelled top, flanked by steep side slopes and a lobate front.

Reference

Rapp, A. (1959) Avalanche boulder tongues in Lappland, *Geografiska Annaler* 41, 34–48.

Further reading

Luckman, B.H. (1978) Geomorphic work of snow avalanches in the Canadian Rockies, *Arctic and Alpine Research* 10, 261–276.

SEE ALSO: hillslope, form; hillslope, process; mass movement; slushflow

BRIAN LUCKMAN

AVALANCHE, SNOW

Controls and characteristics of snow avalanches

In mountainous areas above the snowline, topographical controls result in snow avalanches recurring in locations referred to as avalanche paths. The paths are conventionally divided into the starting zone, track and runout zone. Starting zones occur at elevations above the winter snowline, have slopes in the range of 30° to 45° and are generally lee to the main storm wind directions. Below local treeline elevation, presence of forest cover also influences avalanche location while some secondary topographic factors such as slope form and roughness also play a role. Downslope convexities and transitions from anchoring points to smooth slopes are zones of tension often reflected in avalanche starting points. Across slopes, concavities (bowls and gullies) provide local snow accumulation areas. Tracks and runout zones are at lower angles. The extent of avalanche influence can be determined by their effect on vegetation and contribution to fan-shaped landforms (see Plate 7).

In addition to highest frequencies in winter and spring, avalanche timing is often related to storm events though strong melt may also be significant. Avalanches may be classified in relation to storms as either direct action or climax. The

Plate 7 An avalanche path in the Canadian Rocky Mountains showing a broad bowl-type starting zone, a track devoid of vegetation and a fan-type runout zone

former are initiated by storms (particularly heavy snowfall) and involve only the new snow while the latter owe more to instabilities that develop within the snowpack over periods of at least several days. Climax avalanches may be triggered by new snowfalls but they involve old as well as new snow and are common in spring when the snowpack may contain significant quantities of liquid water.

Snow avalanches are conventionally described in terms of a number of criteria including type of snow, form of motion, snow wetness and depth of failure. The type of snow effects the form of release, such that loose snow avalanches form in cohesionless dry or wet snow, beginning from a point and broadening downslope, while slab avalanches resulting from the existence of a strong layer of snow over a weakly resistant layer, propagate first in a line across the slope. The form of movement after initiation depends on slope steepness, roughness and the nature of the snow. For smooth, relatively gentle slopes and/or wet snow, movement is by flowing in contact with the surface. If the snow is dry and slopes steep and rough, then airborne flow (a powder avalanche) is likely, though most powder avalanches also contain a surface flowing component. Although dry avalanches will generally travel more quickly than wet snow avalanches, wet snow avalanches are capable of transporting considerable quantities of debris and are more likely to be full-depth avalanches. The extreme case of wet avalanches are slush avalanches or SLUSHFLOWS.

Snow avalanches and landforms

Studies of the accumulation of debris transported by avalanches in Scandinavia (Rapp 1960) and Canada (Luckman 1977, 1988; Gardner 1983) show that they may contribute up to several tens of $mm\,a^{-1}$ of accretion on debris slopes with higher rates near the apex. Erosion also occurs

Table 4 Geomorphic effects of snow avalanches

Direct effects				
	Erosional		Depositional	
Bedrock abrasion, transport	Avalanche impact at breaks of slope		Surface sediment redistribution	
Chutes	*Pits and pools*	*Mounds and ramparts* *Debris tails*	*Boulder tongues* *Road-bank*	*Fan*
Indirect effects				
Sediments	Soils	Hydrology, glacier nourishment, snow melt floods	Nivation	

Note: Landforms in italics

though the effects are much more variable and difficult to investigate. While they often occur in association with other processes typical of mountainous areas such as rockfall and DEBRIS FLOWS, in some areas, for example the Himalayas, snow avalanches may be the dominant process in specific elevation zones (Hewitt 1989).

The geomorphic effects of snow avalanches may be either direct or indirect (Table 4). The former may involve both erosional and depositional processes and forms and in the case of some landforms, elements of both. The latter relate to aspects of the environment that influence other geomorphic processes.

The presence of downslope aligned alternating ribs and furrows in rock slopes affected by snow avalanches has long been known (Allix 1924). Abrasion of bedrock by rocks in avalanches is thought to play a role in chute formation but it is probably secondary to the transport of material loosened by other processes (Luckman 1977).

Where snow avalanches occur in locations with significant concave breaks of slope (often where formerly glaciated valleys have been subsequently filled with alluvium), they may generate great impact pressures on the landscape resulting in features collectively referred to as avalanche impact landforms (Luckman *et al.* 1994). The erosional parts of these may form circular or elongated pits but more commonly the pits intersect the local water table or the landforms may occur in water bodies such as lakes or rivers to form pools. They have been described for many areas of the world but seem particularly characteristic in areas of North America, Norway and New Zealand where resistant bedrock has resulted in the preservation of steep-sided formerly glaciated valleys.

Mounds and ramparts are made up of material scooped out by avalanche impacts and often form arcuate ridges at the distal edge of pits or pools. However, there may also be some contribution to their formation by accumulation of debris from frequent small 'dirty' avalanches once the pit or pool is full of previously avalanched snow.

In mountainous areas above the treeline and the winter snowline, snow avalanches redistribute material on debris slopes in association with other processes such as DEBRIS FLOWS and rockfall. Where avalanche transport of debris onto slopes of about 20° to 30° dominates, landforms referred to as avalanche boulder tongues occur. In the pioneering study of these features, Rapp (1959) identified two types – road-bank and fan. Road-banks are smooth flat-topped accumulations of debris often with an asymmetrical profile while fans are fan-shaped tongues of debris extending on relatively low angles towards valley floors. Rapp suggested that fan tongues result from larger avalanches and that the asymmetry of road-bank tongues resulted from preferential deposition of wind-drifted snow on the down valley side of the deposit. However, subsequent studies have suggested that other factors may lead to differences in form with road-bank tongues tending to be favoured where there is a plentiful debris supply and a confined track (Luckman 1977). Boulder tongues are characteristically 100 to 1,000 m long, up to 200 m wide and 10 to 30 m thick. They generally have a strongly concave long profile by which they can be distinguished from debris slopes formed by other processes.

Debris tails are small-scale forms which often occur on boulder tongues where there is a large range in debris sizes. They take the form of streamlined deposits of small to medium-sized particles usually downslope and more rarely upslope of large boulders. As indicated in Table 4, both erosion and deposition may play a role in their formation.

The most significant of the indirect effects of snow avalanches in relation to geomorphology is their influence on the characteristics of sediments and soils, as these may often be used to infer which processes have been responsible for building debris slopes under present or past conditions. Blikra and Nemec (1998) showed that while snow avalanches may transfer surface debris in a similar manner to debris flows, there are significant differences in the resulting sedimentary deposits, including the existence of precariously perched melt-out debris. In addition, snow avalanche deposits are often patchy, ranging from lobes of unsegregated debris to areas with better sorted sands or granules arising from water deposition following snow flows. Snow avalanche deposits may be distinguished from those of rockfalls which are characterized by openwork structure, weaker fabric strength (see FABRIC ANALYSIS) and existence of pronounced downslope increase in sediment size, as shown by Blikra and Nemec (1998) in Norway and Jomelli and Francou (2000) in the French Alps.

References

Allix, A. (1924) Avalanches, *Geographical Review* 14, 519–560.
Blikra, L.H. and Nemec, W. (1998) Post-glacial colluvium in western Norway: depositional processes, facies and paleoclimatic record, *Sedimentology* 45, 909–959.
Gardner, J.S. (1983) Accretion rates on some debris slopes in the Mt. Rae area, Canadian Rocky Mountains, *Earth Surface Processes and Landforms* 6, 347–355.
Hewitt, K. (1989) The altitudinal organisation of Karakoram geomorphic processes and depositional environments, *Zeitschrift für Geomorphologie, NF, Supplementband* 76, 9–32.
Jomelli, V. and Francou, B. (2000) Comparing the characteristics of rockfall talus and snow avalanche landforms in an alpine environment using a new methodological approach; Massif des Ecrins, French Alps, *Geomorphology* 35, 181–192.
Luckman, B.H. (1977) The geomorphic activity of snow avalanches, *Geografiska Annaler* 59A, 31–48.
——(1988) Debris accumulation patterns on talus slopes in Surprise Valley, Alberta, *Géographie physique et Quaternaire* 42, 247–278.
Luckman, B.H., Matthews, J.A., Smith, D.J., McCarroll, D. and McCarthy, D.P. (1994) Snow-avalanche impact landforms – a brief discussion of terminology, *Arctic and Alpine Research* 26, 128–129.
Rapp, A. (1959) Avalanche boulder tongues in Lappland: a description of little-known landforms of periglacial debris accumulation, *Geografiska Annaler* 41, 34–48.
——(1960) Recent development of mountain slopes in Karkevagge and surroundings, northern Scandinavia, *Geografiska Annaler* 42, 73–200.

Further reading

McClung, D. and Schaerer, P. (1993) *The Avalanche Handbook*, Seattle: Mountaineers.

IAN OWENS

AVULSION

Shift of a part or the whole of a river channel to another location on the FLOODPLAIN. It seems to be caused by local superelevation of part of the channel (see CHANNEL, ALLUVIAL) or channel system above the floodplain, as a result of river AGGRADATION, creating a local gradient advantage. Most avulsions occur when a triggering event forces a river across the stability threshold (see THRESHOLD, GEOMORPHIC). The closer the river is to the threshold, the smaller the trigger needed to initiate an avulsion (Jones and Schumm 1999). Local short-term processes triggering avulsions include: tectonic movements, variations in discharge and SEDIMENT LOAD AND YIELD, mass failure, aeolian processes (e.g. the formation of dunes) and log or ice jams. Regional long-term factors controlling avulsion include: BASE LEVEL change, climatic change, tectonic movements and discharge variation (Stouthamer and Berendsen 2000).

References

Jones, L.S. and Schumm, S.A. (1999) Causes of avulsion: an overview, in N.D. Smith and J. Rogers (eds) *Fluvial Sedimentology VI*, Special Publication of the International Association of Sedimentologists 28, 171–178.
Stouthamer, E. and Berendsen, H.J.A. (2000) Factors controlling the Holocene avulsion history of the Rhine–Meuse delta (The Netherlands), *Journal of Sedimentary Research* 70(5), 1,051–1,064.

SEE ALSO: anabranching and anastomosing river; palaeochannel; river delta

ESTHER STOUTHAMER

B

BADLAND

Badlands are deeply dissected erosional landscapes (Plate 8), formed in softrock terrain, commonly but not exclusively in semi-arid regions. Badland processes are dominated by overland-flow erosion. Badlands usually have a high DRAINAGE DENSITY of rill and gully systems, and at most support sparse vegetation. Badlands may comprise zones of coalesced hillslope gullies (see GULLY) within which little of the pre-gullying terrain remains. Badlands are common in areas with at least seasonal drought, in semi-arid and arid areas, Mediterranean and dry-season tropical areas. However, they also occur in humid regions, for example on eroding coastal and river cliffs. Badlands may result from natural processes, but their extent may be accentuated by human activity. Some badlands may be the result of human-induced soil erosion. Two prerequisites for badland development are erodible rock, typically marl, clay, or shale, and available relief. Badlands are common in uplifted and dissected softrock terrain (Plate 8).

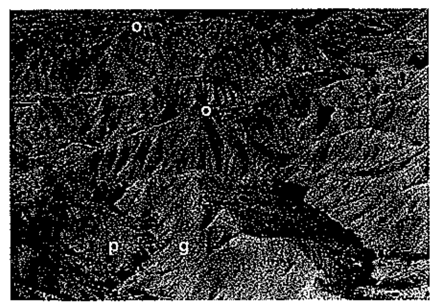

Plate 8 Extensive badland development, Tabernas, south-east Spain. These badlands are cut into older pediment surfaces (o), and owe their origin to tectonically induced dissection of an uplifted Neogene sedimentary basin, under semi-arid climatic conditions. The badland slopes are dominated by surface processes, but note the strong aspect-related contrasts. Note also the differences between micropediment-based badland slopes (p) and gully-based badland slopes (g)

Badland processes and morphology

Processes on badland slopes are dominated by surface erosion by Hortonian (see HORTON'S LAWS) OVERLAND FLOW (Horton 1945), created by rainfall intensity exceeding infiltration capacity. Away from the drainage divide, runoff increases in depth, at first as non-erosive laminar flow, then as turbulent sheet flow. Then, when shear stresses exceed the resistance of the surface, erosion is possible. Initially erosion is by surface winnowing as sheet erosion (see SHEET EROSION, SHEET FLOW, SHEET WASH) or linearly as RILL erosion. As runoff increases further downslope, shear stresses exceed the strength of the underlying material, and channel incision is possible. At that point, defined by a minimum drainage area (Schumm's (1956a) constant of channel maintenance), sheet flow and rill flow give way to open channel flow, and sheet and rill erosion to gully erosion and stream channel processes. The requirements for Hortonian processes are intense storm rainfall, little vegetation cover, low infiltration capacity, easily erodible material and relatively steep slopes.

On slopes dominated by Hortonian processes, smooth rounded divides (Horton's (1945) belt of no erosion) give way downslope to straight rilled slopes, which in turn feed v-shaped gullies at the slope base. Rill and gully networks generally accord with the 'laws' of drainage composition (Schumm 1956a; Strahler 1957) (see HORTON'S LAWS). Local variations in the drainage density of the rill networks or other microtopographic features, such as erosion pinnacles (see HOODOO), reflect local lithological variations in infiltration capacity or erosional resistance.

On many badlands, other processes also operate. Repeated WETTING AND DRYING WEATHERING, and in some areas freezing and thawing, weather the surface materials. These processes have two main effects. Desiccation cracking or frost heave may greatly increase infiltration capacity, so that rainfall-excess overland flow becomes unlikely. In that case surface water infiltrates, and reaches rill systems by lateral flow through the weathered surface layers, reducing interill, but not rill erosion. These processes may be exacerbated by the geochemistry of the material, particularly the presence of exchangeable sodium salts. On wetting, such materials may be prone to slaking, greatly enhancing the potential for mudslides (Benito *et al.* 1993).

Weathering processes, especially on materials rich in swelling clays (see EXPANSIVE SOIL), create their own microtopography of pinnacles (Finlayson *et al.* 1987), crack patterns and so-called 'popcorn textures' (Hodges and Bryan 1982). At a larger scale, the relation between weathering and removal rates (i.e. WEATHERING-LIMITED AND TRANSPORT-LIMITED conditions) may be important. Under transport-limited conditions a weathered mantle may accumulate, ultimately to fail as a shallow mass movement, whereas on an equivalent weathering-limited slope Hortonian processes may dominate. Aspect (see ASPECT AND GEOMORPHOLOGY) often controls these processes (Plate 8), either directly or through its influence on vegetation.

A major process in some badland areas is subsurface erosion by piping (Bryan and Yair 1982; Harvey 1982) (see PIPE AND PIPING; TUNNEL EROSION). Pipes may be induced mechanically by the channelling of overland flow below the surface along animal burrows, vegetation rootways and, particularly in dissected terrain, down tension cracks. Piping is enhanced by the geochemical properties mentioned above (Gutierrez *et al.* 1988; Faulkner *et al.* 2000).

On piped badlands surfaces may have lower rill network densities, and pipe inlets and outlets may add to the morphological diversity. There may be modifications to channel alignments, when the major gullies result from pipe collapse rather than from Hortonian processes.

On some badlands, processes are dominated by single (Hortonian) processes, but on many badlands interactions between several processes take place (Schumm 1956b; Faulkner 1987; Harvey and Calvo-Cases 1991). Spatially, process interactions include on-slope interactions between weathering, Hortonian runoff, mass movements and piping, and also include HILLSLOPE-CHANNEL COUPLING relationships involving interactions between the slope processes and basal stream or gully activity (Harvey 2002). This may involve the build-up of material derived from slope erosion, and its periodic flushing by stream processes, maintaining erosional activity at the slope base (Harvey 1992).

Temporal characteristics of process interactions result from discrepancies between effectiveness, rates and frequency of the various processes. They may relate to individual storm events and recovery periods. However, a common timescale is one of seasonal cyclicity, often related to a seasonal process regime, generating for example seasonal rill development cycles (Schumm 1956a; Harvey 1992). Another type of seasonality or longer-term cyclicity may relate to flushing, when there are different frequencies of sediment generation and removal rates (Harvey 1992; Faulkner 1994). Cyclicity may also be due to discrepancies between weathering and removal rates (Harvey and Calvo-Cases 1991). Over an even longer term there may be progressive changes related to longer-term morphological development (Harvey 1992).

In a wider context, badlands show relationships to GULLY systems. Gully systems may develop as valley-floor (ARROYOs) or hillslope gullies (Campbell 1997). Badlands result from the coalescence of both basally- and midslope-induced hillslope gullies. Of fundamental importance for badland morphology and development is the local base level. An incising or laterally migrating gully channel maintains an active base level, influencing all badland processes, surface processes, slope stability, sediment removal, and even subsurface processes through its influence on hydraulic gradients. Basally-induced gullying and badland development are more likely to have effective base-level control, but even there slope retreat may transform gully/channel-based badlands by micropediment-based badlands (Plate 8).

Badland dynamics

Badlands have two major roles within the context of the wider geomorphic system: (1) as major sediment sources to the fluvial system, and (2) as a major influence on slope evolution. Badlands, especially when coupled with the stream network, represent a zone of drainage network expansion, an increase in drainage density and an increase in stream order (Strahler 1957). This has hydrological implications, but above all increases the sediment supply to the fluvial system to the extent that a zone of badlands may dominate sediment dynamics of a drainage basin (Campbell 1997). Badlands may act as a major influence on slope evolution, especially in semi-arid areas, producing extensive pediment areas at the base of retreating escarpments.

Badlands are erosional not equilibrium forms. In addition to the results of process interactions, badland morphology progressively changes as the badlands develop. This, in turn, modifies the processes. Ultimately badland development depends on the relative rates of extension and stabilization. Harvey (1992) has demonstrated that in a humid environment, once a gully system is decoupled from a basal stream, stabilization by vegetation colonization operates faster than gully extension. Those gullies do not develop into badlands. However, in many semi-arid areas, although auto-stabilization mechanisms have been recognized (Alexander *et al.* 1994), stabilization processes are slower than gully extension so that badlands develop and persist. Under conditions of incising base levels, basal v-shaped gully systems would maintain characteristic badland processes and morphology on the slopes. Under stable base-level conditions the badland slopes progressively retreat, forming pediment-based badlands, which if they do not self-stabilize, would ultimately produce a landscape of extensive pediments and small badland residuals.

One factor of fundamental importance is the interaction between vegetation and geomorphic processes which affects both the generation of overland flow and the stabilization of eroded slopes (Alexander *et al.* 1994; Gallert *et al.* 2002). However, of the main factors influencing badland geomorphology, it is the interaction of base level with the surface processes that has the greatest influence on badland evolution.

References

Alexander, R.W., Harvey, A.M., Calvo A., James, P.A. and Cerda, A. (1994) Natural stabilisation mechanisms on Badland slopes, Tabernas, Almeria, Spain, in A.C. Millington and K. Pye (eds) *Environmental Change in Drylands: Biogeographical and Geomorphological Perspectives*, 85–111, Chichester: Wiley.

Benito, G., Gutierrez, M. and Sancho, C. (1993) The influence of physico-chemical properties on erosion processes in badland areas, Ebro basin, NE Spain, *Zeitschrift für Geomorphologie* 37, 199–214.

Bryan, R. and Yair, A. (eds) (1982) *Badland Geomorphology and Piping*, Norwich: Geobooks.

Campbell, I.A. (1997) Badlands and badland gullies, in D.S.G. Thomas (ed.) *Arid-Zone Geomorphology*, 2nd edition, 261–291, Chichester: Wiley.

Faulkner, H. (1987) Gully evolution in response to both snowmelt and flash flood erosion, western Colorado, in V. Gardiner (ed.) *International Geomorphology*, vol. 1, 947–969, Chichester: Wiley.

——(1994) Spatial and temporal variations of sediment processes in the alpine semi-arid basin of Alkali Creek, Colorado, USA, *Catena* 9, 203–222.

Faulkner, H., Spivey, D. and Alexander, R.W. (2000) The role of some site geochemical processes in the development and stabilisation of three badland sites in Almeria, *Geomorphology* 35, 87–99.

Finlayson, B.L., Gerits, J.J.P. and van Wesermael, B. (1987) Crusted microtopography on badland slopes in southeast Spain, *Catena* 14, 131–144.

Gallert, F., Sole, A., Puigdefabregas, J. and Lazaro, R. (2002) Badland systems in the Mediterranean, in L.J. Bull and M.J. Kirkby (eds) *Dryland Rivers: Hydrology and Geomorphology of Semi-arid Channels*, 299–326, Chichester: Wiley.

Gutierrez, M., Benito, G. and Rodriguez, J. (1988) Piping in badland areas of the middle Ebro basin, Spain, *Catena Supplement* 13, 49–60.

Harvey, A.M. (1982) The role of piping in the development of badlands and gully systems in south east Spain, in R. Bryan and A. Yair (eds) *Badland Geomorphology and Piping*, 317–335, Norwich: Geobooks.

——(1992) Process interactions, temporal scales and the development of hillslope gully systems: Howgill Fells, northwest England, *Geomorphology* 5, 323–344.

——(2002) Effective timescales of coupling within fluvial systems, *Geomorphology* 44, 175–201.

Harvey, A.M. and Calvo-Cases, A. (1991) Process interactions and rill development on badland and gully slopes, *Zeitschrift für Geomorphologie, Supplementband* 83, 175–194.

Hodges, W.K. and Bryan, R.B. (1982) The influence of material behaviour on runoff initiation in the Dinosaur Badlands, Canada, in R. Bryan and A. Yair (eds) *Badland Geomorphology and Piping*, 13–46, Norwich: Geobooks.

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.

Schumm, S.A. (1956a) Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey, *Geological Society of America Bulletin* 67, 597–646.
——(1956b) The role of creep and rainwash on the retreat of Badland slopes, *American Journal of Science* 254, 693–700.
Strahler, A.N. (1957) Quantitative analysis of watershed geomorphology, *American Geophysical Union, Transactions* 38, 913–920.

ADRIAN HARVEY

BAJADA

The broad zone of coalesced or compound ALLUVIAL FANs that form a more or less continuous piedmont alluvial apron lying between the mountain front and the basin floor in areas like the semi-arid south western United States. They are in contrast to rock-cut PEDIMENTs. The term was introduced by Tolman (1909).

Reference

Tolman, C.F. (1909) Erosion and deposition in the southern Arizona bolson region, *Journal of Geology* 17, 136–163.

A.S. GOUDIE

BANK EROSION

Bank erosion is the detachment and entrainment of bank material as grains, aggregates or blocks by fluvial, subaerial or geotechnical processes. Riverbank erosion processes are important for the evolution of MEANDERING and BRAIDED RIVER systems and FLOODPLAINs, catchment sediment output and biodiversity on floodplains. Bank erosion can also lead to loss of agricultural land and riparian structures, exacerbated sedimentation problems and riverine boundary disputes, sometimes necessitating bank stabilization works.

Bank erosion measurement

The many methods of bank retreat measurement can be classified into long-timescale, medium-timescale, and short-timescale techniques (Lawler 1993a). Long-timescale methods employ sedimentological, botanical or cartographic evidence to reveal channel change over decades or thousands of years. For example, sequences of channel movements can be preserved in datable fluvial deposits, or quantified through dendrochronological analysis (see DENDROCHRONOLOGY) of trees colonizing bar (see BAR, RIVER) surfaces (e.g. Hickin and Nanson 1984). River course changes can be quantified by superimposing early maps, aerial photographs and satellite imagery (e.g. Hooke and Redmond 1989; Lewin 1987), often using analytical photogrammetry and GIS (e.g. Lane *et al.* 1993). Medium-timescale techniques include the periodic field resurvey of bank lines with theodolites, EDMs (Electronic Distance Measurers), Total Stations or GPS (Global Positioning Systems) (Lawler 1993a). Cross-section resurvey, however, using levelling or Total Station techniques, can be more sensitive to subtle changes. Airborne laser altimetry and side-scan sonar methods have also been applied.

The following short-timescale techniques are more useful for process studies, because the geomorphological change can be related to forcing hydrological and meteorological events. Erosion pins can be inserted into banks: erosion then exposes more pin, the length of which is recorded periodically (Lawler 1993a). Terrestrial photogrammetric monitoring involves the repeated capture of ground photograms using stereometric cameras, from which the three-dimensional bank form (DIGITAL ELEVATION MODEL (DEM)) is derived. 'Subtracting' successive DEMs reveals the intervening bank erosion rate (Lawler 1993a; Lane *et al.* 1993). However, all the methods above reveal little about the *timing* of bank erosion *events*, knowledge which is crucial to process inference. Lawler (1992), therefore, developed the *automatic* Photo-Electronic Erosion Pin (PEEP) system. When erosion occurs, the PEEP signal increases; if deposition occurs, voltages are decreased. The system thus allows the magnitude, timing and frequency of erosional and depositional activity to be monitored more precisely, including for TIDAL CREEK (Lawler *et al.* 2001), and is now used by twenty research groups worldwide. The example in Figure 9 shows how the PEEP approach, for the first time, fixes the time of an erosion event to forty-three hours *after* the hydrograph peak; this suggests the operation of mass failure processes rather than fluid entrainment.

Bank erosion rates

Bank erosion rates range from 0–1,000 $m\,a^{-1}$ and tend to increase with boundary shear stress, STREAM POWER, FREEZE–THAW CYCLE activity and for silty or saturated bank materials of low cohesion (Lawler *et al.* 1997). Within some basins bank erosion rates peak in the *middle* reaches (e.g. Lewin 1987; Lawler *et al.* 1999), possibly related to stream power increases (Lawler 1992; Abernethy and Rutherfurd, 1998; Knighton, 1999).

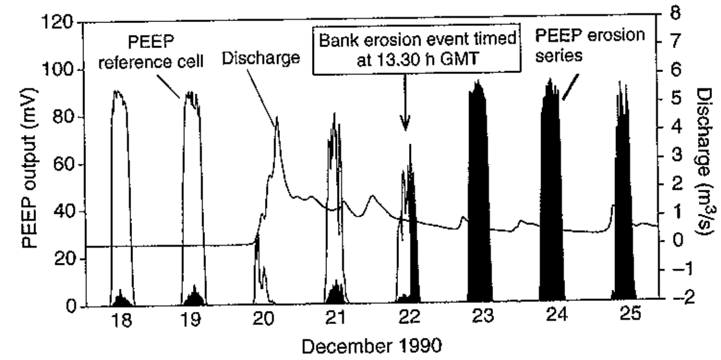

Figure 9 Timing of a bank erosion event at 13.30h GMT on 22 December 1990 detected by the automatic Photo-Electronic Erosion Pin (PEEP) monitoring system on the Upper River Severn, UK. The bank erosion event lags the flow peak by forty-three hours, suggesting mass failure processes are responsible (from Lawler *et al.* 1997)

Riverbank erosion processes

The many bank erosion mechanisms identified can be grouped into fluid entrainment, preparation or mass failure processes (Lawler 1992; Lawler *et al.* 1997; Prosser *et al.* 2000).

FLUID ENTRAINMENT PROCESSES

Entrainment occurs when the motivating forces due to the flow (lift and drag, often indexed by boundary shear stress) and particle mass exceed the friction and cohesion forces holding the particle in place (Lawler *et al.* 1997). On non-cohesive (e.g. sandy) banks, material is usually entrained grain by grain, while on cohesive, fine-grained banks, material is eroded as aggregates or crumbs bound by cohesive forces. Cohesion results from a combination of physico-chemical, inter-granular, bonding forces, driven by the mineralogy, dispersivity, moisture content and particle size distribution of the bank material, and the temperature, pH and electrical conductivities of the pore fluid and river water (Osman and Thorne 1988). Cohesive banks are normally much more resistant to fluid entrainment than non-cohesive ones.

PREPARATION PROCESSES

The ERODIBILITY of cohesive bank materials, however, can change because of preparation or weakening processes. Crucially, then, the critical shear stress value for entrainment varies with antecedent material preparation (Lawler 1992; Prosser *et al.* 2000). For example, desiccated banks may crack as moisture is thermally driven off and clay minerals shrink (Plate 9). For instance, in summer, east-facing banks of the river Arrow, Warwickshire, UK reach early-morning warming rates of 7 $°C\,h^{-1}$, peak daily temperatures above 30 °C and diurnal temperature ranges of 20 °C (Lawler 1992). Flowing waters then exploit cracks to enhance erosion (e.g. Lawler 1992). Freeze–thaw activity takes many forms (e.g. Lawler 1993b; Prosser *et al.* 2000). For example, NEEDLE-ICE can lift or incorporate material, and transport it downslope on ablation. Much disturbed sediment remains, though, to be readily removed by subsequent flow rises (Lawler 1993b).

MASS FAILURE PROCESSES

Mass failure occurs when blocks of material collapse or slide towards the bank toe (Plate 9). Banks are vulnerable to mass failure if steep, high, fine-grained, of high bulk unit weight and subject to high or fluctuating PORE-WATER PRESSURES – indeed any variable which increases the mass of material above a potential failure

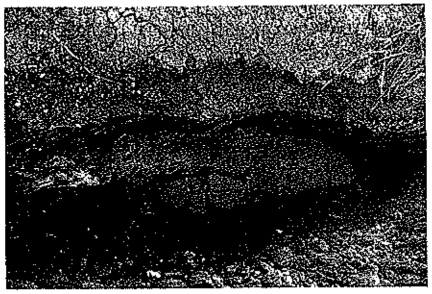

Plate 9 Bank erosion on the lowland river Arrow, Warwickshire, UK. A failed block of bank material (length ~2 m) lies under water at the bank toe, with the scar of the failure surface visible above. Desiccation cracking is above the scar. Flow is from right to left

surface. Hence, bed scour can induce bank failure by increasing bank height and angle. Also, failure can follow increases in block mass due to moisture uptake, often after submergence. Bank failures should thus occur on hydrograph recession limbs, following saturation. This is confirmed by the PEEP automatic erosion monitoring system (Figure 9). Figure 10 shows bank failure characteristics. For the most common type, the failure surface is almost planar (Plate 9 and Figure 10). Cantilever failure occurs on composite riverbanks if, because of faster erosion of the lower more erodible layers, an overhang develops then collapses (Lawler *et al.* 1997) (Figure 10g and h).

Mass failures can be analysed using geotechnical slope stability theory (e.g. Osman and Thorne 1988; Darby and Thorne 1996; the CONCEPTS model of the United States Department of Agriculture (USDA)). One example is the Culmann formula for planar failure (Figure 10c), which predicts the critical height for a bank, H_{crit}:

$$H_{crit} = \frac{4c}{\gamma} \cdot \frac{\sin\alpha \cdot \cos\phi}{\{1-\cos(\alpha-\phi)\}}$$

where c = material cohesion (kPa), γ = *in situ* unit weight of material ($kN\,m^{-3}$), α = slope angle (°), ϕ = friction angle (°). Many of the data required can be collected using the Stream Reconnaissance Record Sheets of Thorne *et al.* (1996).

Bank processes may change in a longitudinal direction (Lawler 1992). This idea, developed into the DOCPROBE model (DOwnstream Change in the PROcesses of Bank Erosion), suggests that, in upstream reaches, preparation processes are most effective, because stream power and bank heights are too low for significant fluid entrainment and mass failure respectively. In middle reaches, where stream power is high, fluid entrainment dominates. Further downstream, bank heights and material properties exceed critical values and mass failure processes prevail. Evidence has now emerged to support this model (e.g. Lawler 1992; Knighton 1999; Abernethy and Rutherfurd 1998). Vegetation can considerably reduce fluid erosion, partly through root reinforcement or foliage protection (Thorne 1990; Abernethy and Rutherfurd 1998; Simon and Collinson 2002). However, forest canopy shade may suppress shorter riparian vegetation, increasing erosion.

A much richer mix of bank erosion processes is now recognized. Though flow processes are important, research has shifted to temporal change in bank erodibility, vegetation, riparian hydrology and the dynamics of bank erosion events, often using novel automated monitoring techniques.

References

Abernethy, B. and Rutherfurd, I.D. (1998) Where along a river's length will vegetation most effectively stabilize stream banks? *Geomorphology* 23, 55–75.

Darby, S.E. and Thorne, C.R. (1996) Development and testing of riverbank-stability analysis, *Proceedings of the American Society of Civil Engineers, Journal of Hydraulic Engineering* 122, 443–454.

Hey, R.D., Heritage, G.L., Tovey, N.K., Boar, R.R., Grant, N. and Turner, R.K. (1991) Streambank protection in England and Wales, Research and Development Note 22, London: National Rivers Authority.

Hickin, E.J. and Nanson, G.C. (1984) Lateral migration of river bends, *Proceedings of the American Society of Civil Engineers, Journal of Hydraulic Engineering* 110, 1,557–1,567.

Hooke, J.M. and Redmond, C.E. (1989) River-channel changes in England and Wales, *Journal of the Institute of Water Engineers and Managers* 3, 328–335.

Knighton, A.D. (1999) Downstream variation in stream power, *Geomorphology* 29, 293–306.

Lane, S.N., Richards, K.S. and Chandler, J.H. (1993) Developments in photogrammetry; the geomorphological potential, *Progress in Physical Geography* 17, 306–328.

Lawler, D.M. (1992) Process dominance in bank erosion systems, in P. Carling and G.E. Petts (eds)

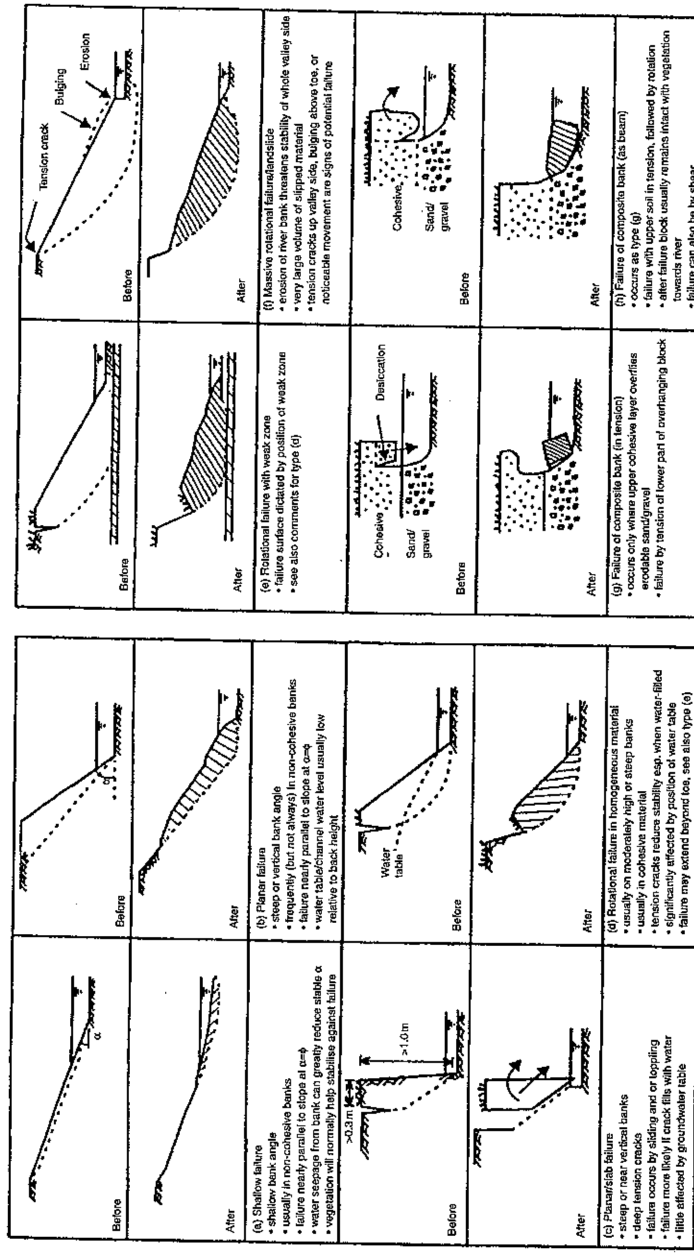

Figure 10 Characteristics of bank failure (from Hey *et al.* 1991; cited in Lawler *et al.* 1997)

Lowland Floodplain Rivers: Geomorphological Perspectives, 117–143, Chichester: Wiley.
Lawler, D.M. (1993a) The measurement of river bank erosion and lateral channel change: a review, *Earth Surface Processes and Landforms* 18, 777–821.
——(1993b) Needle-ice processes and sediment mobilization on river banks: the River Ilston, West Glamorgan, UK, *Journal of Hydrology* 150, 81–114.
Lawler, D.M., Grove, J., Couperthwaite, J.S. and Leeks, G.J.L. (1999) Downstream change in river bank erosion rates in the Swale–Ouse system, northern England, *Hydrological Processes* 13, 977–992.
Lawler, D.M., Thorne, C.R. and Hooke, J.M. (1997) Bank erosion and instability, in C.R. Thorne, R.D. Hey and M.D. Newson (eds) *Applied Fluvial Geomorphology for River Engineering and Management*, 137–172, Chichester: Wiley.
Lawler, D.M., West, J.R., Couperthwaite, J.S. and Mitchell, S.B. (2001) Application of a novel automatic erosion and deposition monitoring system at a channel bank site on the tidal River Trent, UK, *Estuarine, Coastal and Shelf Science* 53, 237–247.
Lewin, J. (1987) Historical river channel changes, in K.J. Gregory, J. Lewin and J.B. Thornes (eds) *Palaeohydrology in Practice*, 161–175, Chichester: Wiley.
Osman, A.M. and Thorne, C.R. (1988) Riverbank stability analysis. I: Theory, *Proceedings of the American Society of Civil Engineers Journal of Hydraulic Engineering* 114, 134–150.
Prosser, I.P., Hughes, A.O. and Rutherfurd, I.D. (2000) Bank erosion of an incised upland channel by subaerial processes: Tasmania, Australia, *Earth Surface Processes and Landforms* 25, 1,085–1,101.
Simon, A. and Collinson, A.J.C. (2002) Quantifying the mechanical and hydrologic effects of riparian vegetation on streambank stability, *Earth Surface Processes and Landforms* 27, 527–546.
Thorne, C.R. (1990) Effects of vegetation on riverbank erosion and stability, in J.B. Thornes (ed.) *Vegetation and Erosion*, 125–144, Chichester: Wiley.
Thorne, C.R., Allen, R.G. and Simon, A. (1996) Geomorphological river channel reconnaissance for river analysis, engineering and management, *Transactions of the Institute of British Geographers* 21, 469–483.
USDA CONCEPTS model at: http://msa.ars.usda.gov/ms/oxford/nsl/agnps/concepts/concepts_dl.html

SEE ALSO: fluvial geomorphology; hydraulic geometry

DAMIAN LAWLER

BANKFULL DISCHARGE

Bankfull discharge, a hydrologic term, is the flow rate when the stage (height) of a stream is coincident with the uppermost level of the banks – the water level at channel capacity, or bankfull stage. Bankfull stage is a fluvial-geomorphic term (see FLUVIAL GEOMORPHOLOGY) requiring an interpretation of site-specific landforms. In this context, bank typically refers to a sloping margin of a natural, stream-formed, alluvial channel (see CHANNEL, ALLUVIAL) that confines discharge during non-flood flow. Although the term bankfull stage can refer to various channel-bank levels, it generally applies to alluvial-stream channels (1) having sizes and shapes adjusted to recent fluxes of water and sediment, (2) that are principal conduits for discharges moving through a length of alluvial bottomland, and (3) that are bounded by FLOODPLAINS upon which water and sediment spill when the flow rate exceeds that of bankfull discharge. Thus, the concept of bankfull discharge, which often approximates the mean annual flood for perennial streams, includes the floodplain as a unique, identifiable geomorphic surface, all higher surfaces of alluvial bottomlands being terraces (see TERRACE, RIVER) (generally former floodplain surfaces), and acknowledgement that bankfull discharge occurs only when stream stage is at floodplain level.

Previous studies

Numerous discussions, only a few of which are cited here, have addressed the bankfull concept. A review by Williams (1978) documented a variety of published definitions for bankfull discharge and listed a wide range of flood frequencies related to the bankfull stage; almost without doubt the range resulted from observer misidentification of surfaces higher and lower than the floodplain as that of bankfull. Radecki-Pawlik (2002), among many others, showed that field determinations of bankfull stage and therefore floodplain level are interpretive, and thus the related bankfull discharge may differ greatly from that of the mean annual flood.

Papers by Woodyer (1968) and Osterkamp and Hupp (1984) recognized a variety of bottomland surfaces, including the floodplain, and related an approximate flow characteristic to each. Petit and Pauquet (1997) and Castro and Jackson (2001), respectively, determined return intervals for bankfull discharge of generally 1.2 to 3.3 years at sites of gravel-bed streams in France, and return intervals of 1.0 to 3.11 years for bankfull discharge at seventy-five stream sites of the north-western United States.

Significance

Bankfull discharge is significant owing to the hydraulic and related physical changes that occur for most alluvial stream channels when flow increases from in-channel to overbank conditions. Resistance to flow 'decreases with increasing water depth to reach a minimum at bankfull stage, so that the channel operates most efficiently with regard to water conveyance when the flow is at bankfull level' (Petts and Foster 1985: 150). Thus, the change in hydraulics as flow depth increases exerts a basic control on the geomorphic processes that are related to floodplain formation, regardless of the return period that may be associated empirically with the discharge at bankfull stage. The approximate height above the channel bed at which overbank flow begins is the level to which riverine bars develop (see BAR, RIVER), a process that can be described in terms of flow field, channel bathymetry and characteristics of sediment size and transport (Nelson and Smith 1989). Significant also is the observation that numerous data collected from perennial streams suggest that floodplain level is roughly equivalent to the stage of the mean annual flood, about 2.3 years (Wolman and Leopold 1957).

Floodplain formation

Floodplains typically form through POINT BAR deposition and, generally to a lesser extent, by deposition of sediment during overbank flows. Overbank deposits underlying floodplain and alluvial-terrace surfaces are typically poorly sorted and generally exhibit thinly bedded and alternating layers of silt, sand and possibly gravel. When a succession of floods causes overbank deposition, each flood elevating the surface higher above the channel, the deposits tend to grade from relatively coarse particles at the bottom upward to finer sizes. Because the thickness of overbank sediment deposited by large floods is generally small, averaging about 20 mm (Wolman and Leopold 1957), numerous episodes of overbank deposition are ordinarily needed to result in the accumulation of sediment on a gravel bar to a flood-plain level. AGGRADATION above flood-plain level is minimal owing to the infrequency of overtopping discharges, scour of flood-plain sediment by large floods and EROSION of accumulated deposits by lateral channel migration (Wolman and Leopold 1957). Nanson (1986), among others, suggested that processes resulting in the development of floodplains are influenced by prevailing conditions of energy (largely channel gradient) and the availability of sediment for entrainment. It follows, therefore, that many high-gradient streams have little potential for lateral ACCRETION (point-bar formation) and that flood-plain development occurs principally by vertical accretion.

Considerations and problems

The legitimate application of bankfull stage, bankfull discharge and floodplain to hydrologic and geomorphic studies requires accurate field determination of bankfull (flood-plain) level, which may be difficult if streamflow data are unavailable. Bankfull level is recognized easily along channels with point-bar deposits, especially if recent overbank deposits overlie point-bar sediment. For channels lacking point-bar features, interpretation of bankfull stage must rely on observations of channel morphology and gradient, bed and bank sediment, vegetation, root exposure and indications of flood processes.

The bankfull concept has been valuable as a means of describing clearly and effectively the processes and landforms of perennial-stream channels in humid areas. In recent decades, however, the bankfull concept has been so prevalent that it often has been overextended by Earth scientists who have confused the floodplain with other prominent alluvial landforms and corresponding flow frequencies. Thus, 'bankfull' should be used cautiously and consistently within the constraints by which it was described.

Common difficulties of the bankfull concept include its misapplication to non-alluvial conditions, such as streams incising till or debris-flow deposits, where fluvial adjustment is incomplete and bank height may be poorly related to flood frequency. Alluvial surfaces adjacent to high-energy (especially alpine and subalpine) streams, which typically correspond to the approximate stage of mean discharge (Osterkamp and Hupp 1984; Hupp 1986), commonly are misidentified as floodplain. As noted previously, the approximate correlation of bankfull stage (flood-plain level) and discharge with the mean annual flood pertains principally to perennial streams of moist areas; in drier regions with intermittent to highly ephemeral streams, the floodplain may correspond to floods with return periods of 100 years or more.

Flows smaller than bankfull discharge, those that occur more frequently than that of bankfull stage, typically cause in-channel features unrelated to bankfull discharge (such as BEDFORMS and bars). Processes and channel features resulting from these common events should not be confused with processes that correspond to bankfull discharge, and it should be recognized that all flows transport and sort sediment, thereby modifying the stream-channel morphology. Inappropriate emphasis on the bankfull concept has given rise to terms such as 'dominant discharge' and 'channel-forming discharge'. Such terminology, which focuses on a single flow rate, fails to recognize that all flows contribute to channel shape. As demonstrated by Wolman and Miller (1960), bankfull discharge, if related to geomorphic work accomplished, may indeed be dominant when applied to well-adjusted channels of perennial streams, but the dominance of a bankfull discharge becomes increasingly questionable as rates of precipitation and runoff decrease. Because all flows alter the shape of alluvial-stream channels, the term 'channel-forming discharge' may be inappropriate.

References

Castro, J.N. and Jackson, P.L. (2001) Bankfull discharge recurrence intervals and regional hydraulic geometry relationships: patterns in the Pacific Northwest, USA, *Journal of the American Water Resources Association* 37, 1,249–1,262.

Hupp, C.R. (1986) The headward extent of fluvial landforms and associated vegetation on Massanutten Mountain, Virginia, *Earth Surface Processes and Landforms* 11, 545–555.

Nanson, G.C. (1986) Episodes of vertical accretion and catastrophic stripping: a model of disequilibrium flood-plain development, *Geological Society of America Bulletin* 97, 1,467–1,475.

Nelson, J.M. and Smith, J.D. (1989) Evolution of erodible channel beds, in S. Ikeda and G. Parker (eds) *River Meandering*, 321–77, Washington, DC: AGU Water Resources Monograph 12.

Osterkamp, W.R. and Hupp, C.R. (1984) Geomorphic and vegetative characteristics along three northern Virginia streams, *Geological Society of America Bulletin* 95, 1,093–1,101.

Petit, F. and Pauqet, A. (1997) Bankfull discharge recurrence interval in gravel-bed rivers, *Earth Surface Processes and Landforms* 22, 685–693.

Petts, Geoff and Foster, Ian (1985) *Rivers and Landscape*, London: Edward Arnold.

Radecki-Pawlik, Artur (2002) Bankfull discharge in mountain streams: theory and practice, *Earth Surface Processes and Landforms* 27, 115–123.

Williams, G.P. (1978) Bank-full discharge of rivers, *Water Resources Research* 14, 1,141–1,154.

Wolman, M.G. and Leopold, L.B. (1957) *River Flood Plains: Some Observations on their Formation*, Washington, DC: US Geological Survey Professional Paper 282-C.

Wolman, M.G. and Miller, J.P. (1960) Magnitude and frequency of forces in geomorphic processes, *Journal of Geology* 68, 54–74.

Woodyer, K.D. (1968) Bankfull frequency in rivers, *Journal of Hydrology* 6, 114–142.

W.R. OSTERKAMP

BAR, COASTAL

Coastal bars can broadly be defined as aggradational ridges of sediments whose formation, morphology and behaviour are determined by interactions between WAVES, CURRENTS, tides, local slope and grain size. Some confusion exists regarding usage of the terms *bar* and *ridge*, however, since features such as BEACH RIDGES and CHENIER RIDGES are not normally considered bars. Furthermore, bars are found in BEACH, RIVER DELTA, ESTUARY and CONTINENTAL SHELF environments with a wide range of sizes, types and orientation. However, comparatively more COASTAL GEOMORPHOLOGY research studies have focused on bars which exist in the nearshore zone of sandy wave-dominated beaches.

Early studies (e.g. Shepard 1950) identified a seasonal cycle of beach morphology with winter storms promoting offshore sediment transport and bar formation and calmer conditions in summer favouring landward migration of the bar and eventual welding to the beach face. However, the existence of such 'winter' and 'summer' profiles is not universal as both barred and non-barred profiles occur at all times in some areas, while in others only one type may persist throughout the year. Furthermore, cyclic beach response at timescales much shorter than seasons can result in barred/non-barred profiles (Short 1979).

Cross-shore barred profiles are most commonly asymmetrical, having a distinct crest and a steeper landward slope than seaward slope (Figure 11a). Types of bars are often distinguished based on their alongshore planform shape and orientation relative to the shoreline. They may be linear (also referred to as longshore or shore-parallel; see Figure 11b), sinuous or crescentic (often termed

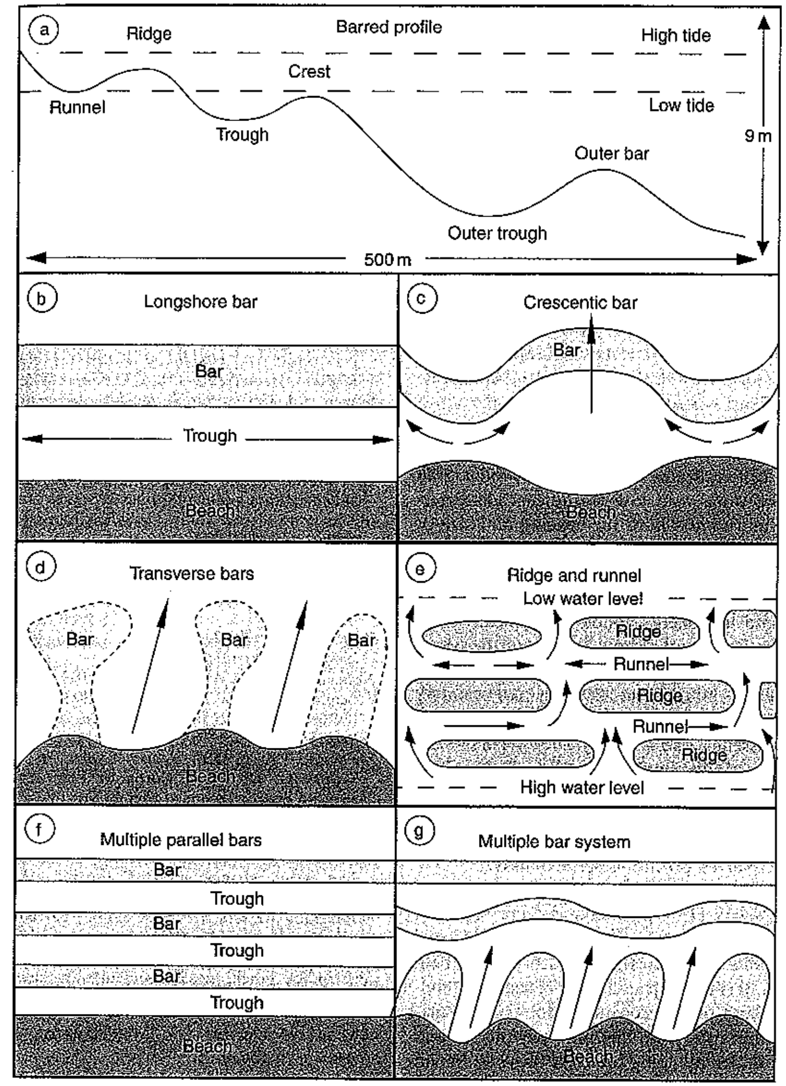

Figure 11 Idealized cross-shore barred profile (a) and some examples of bar types (b–g)

rhythmic) with a trough separating them from the shoreline (Figure 11c), or they may consist of alternating transverse bars, which are welded to the shoreline and are separated by channels occupied by RIP CURRENTS (Figure 11d). Bar type appears to be strongly related to wave energy level with linear bars developing under high-energy conditions, crescentic bars during intermediate energy, and transverse bars during lower wave energy levels. Under very low-energy conditions, a bar may become fully welded to the beach and appear as a flat terrace at low tide. These types of bar configurations are common on microtidal beaches and may grade into each other as energy levels vary. A number of classifications exist describing both bar types and the continuum of bar evolution (e.g. Greenwood and Davidson-Arnott 1979; Short and Aagaard 1993; Wijnberg and Kroon 2002).

Many beaches are characterized by the existence of multiple offshore bars (Figure 11f, g), ranging in number from two to over a dozen in some cases. Although high-energy swell wave environments can exhibit two or three bars, multi-barred profiles are most commonly developed in storm wave dominated sea and lacustrine environments where the outer bars are only active during short intense storms, remaining stationary during longer periods of low-wave energy. The number of bars seems to be related to the overall nearshore slope with lower gradients characterized by more bars. Both the spacing and size of bars has been observed to increase offshore. Bars are commonly absent on steep beaches.

Sandy beaches characterized by significant tidal ranges and low-energy conditions typically have gentle gradients and although some of the previously described bar types may be present, the presence of RIDGE AND RUNNEL TOPOGRAPHY (Figure 11e) in the intertidal zone is more common (Masselink and Anthony 2001). These exist as a series of low amplitude bars (ridges), which are usually stable in form and position, separated by subdued channels (runnels) associated with tidal drainage, and should not be confused with the multi-barred profiles described above.

As reviewed by Komar (1998) and Aagaard and Masselink (1999), some uncertainty remains regarding the formation of nearshore bars despite considerable theoretical, laboratory and field research. Bars develop as a result of sediment convergence and most mechanisms for their formation attempt to explain this. An early theory, that vortices under plunging breakers move sediment seaward forming a bar just seaward of the breakpoint, has largely been discounted since the breakpoint location on natural beaches, with irregular waves varies considerably. It is more likely that sediment convergence results from onshore sediment transport outside the surf zone due to wave asymmetry and offshore transport in the surf zone due to bed return flow, with the bar forming somewhere near the breakpoint. Single and multiple bar formation has also been attributed to net sediment transport patterns associated with standing infragravity waves. According to this theory, if the sediment transport is predominantly bedload the bars will form at nodal positions, whereas if suspension dominates, they will form at antinodal positions (Bowen 1980). This theory has also proven useful in explaining rhythmic bar morphology (Holman and Bowen 1982) although nearshore cell circulation and rip currents are also important factors.

References

Aagaard, T. and Masselink, G. (1999) The surf zone, in A.D. Short (ed.) *Handbook of Beach and Shoreface Morphodynamics*, 72–118, Chichester: Wiley.

Bowen, A.J. (1980) Simple models of nearshore sedimentation; beach profiles and longshore bars, in S.B. McCann (ed.) *The Coastline of Canada*, 1–11, Geological Survey of Canada, Paper 80–10.

Greenwood, B. and Davidson-Arnott, R.G.D. (1979) Sedimentation and equilibrium in wave-formed bars: a review and case study, *Canadian Journal of Earth Sciences* 18, 424–433.

Holman, R.A. and Bowen, A.J. (1982) Bars, bumps and holes: models for the generation of complex beach topography, *Journal of Geophysical Research* 87, 457–468.

Komar, P.D. (1998) *Beach Processes and Sedimentation*, New Jersey: Prentice Hall.

Masselink, G. and Anthony, E.J. (2001) Location and height of intertidal bars on macrotidal ridge and runnel beaches, *Earth Surface Processes and Landforms* 26, 759–774.

Shepard, F.P. (1950) *Beach Cycles in Southern California*, US Army Corps of Engineers, Beach Erosion Board, Technical Memo No. 15.

Short, A.D. (1979) Three dimensional beach-stage model, *Journal of Geology* 87, 553–571.

Short, A.D. and Aagaard, T. (1993) Single and multi-bar beach change models, *Journal of Coastal Research*, Special Issue 15, 141–157.

Wijnberg, K.M. and Kroon, A. (2002) Barred beaches, *Geomorphology* 48, 103–120.

SEE ALSO: beach; beach sediment transport; current; ridge and runnel topography; wave

ROBERT W. BRANDER

BAR, RIVER

The nature and distribution of alluvial instream geomorphic units is fashioned by the interaction between unit stream power along a river reach and sediment calibre and availability. If a reach has excess energy relative to available sediment of sufficient size, flushing is likely to occur. Alternatively, with excess sediment availability or insufficient flow energy, continuous instream sedimentation is likely to occur, commonly in the form of near-homogenous sheets. In most cases, rivers fall somewhere between these extremes, with transient sediment stores of differing calibre and bed material organization in differing landforms along the channel bed.

The most common instream geomorphic units are accumulations of deposits referred to as bars. These areas of net sedimentation of comparable size to the channels in which they occur are key indicators of within-channel processes. Interpretation of bar type is often critical in elucidation of river character and behaviour. There are two main components in bar form. The basal feature, or platform, is made up of coarse material and is overlain by supraplatform deposits of varying forms which is subject to removal and replacement during floods. Bars are readily reworked as channels shift position over the valley floor. Bank-attached features are much less likely to be reworked than mid-channel forms. The long-term preservation of bars is conditioned by factors such as the aggradational regime and the manner of channel movement.

Bars adopt many varied morphologies, ranging from simple unit bars (Smith 1970) to complex compound features (Brierley 1991, 1996). Bar character is controlled primarily by local-scale flow and grain size characteristics. Unit bars are simple features composed of one depositional style. The sediments of a unit bar (whether they be sand or gravel in composition) tend to fine in a downstream direction. As unit bars are found at characteristic locations along long profiles under particular sets of flow energy (stream power) and bed material texture relationships, a 'typical' down-valley transition in forms can be discerned (Church and Jones 1982). Bed material character, and the competence of flow to transport it, determine formation of longitudinal bars as flow divides around a tear-drop shaped feature. When flow is oriented obliquely to the long axis of the bar, a diagonal feature is produced. This is commonly associated with a dissected riffle. In highly sediment-charged sandy conditions, flow divergence results in transverse or linguoid bars, which extend across rather than down the channel (Collinson 1970; Cant and Walker 1978). Alternatively, the entire channel bed may comprise a homogenous sandsheet.

Instances in which patterns of sedimentation are dominated by within-channel bars reflect situations in which the material on the channel bed is either too coarse to be transported or the volume of material is too great to be transported. These scenarios are generally associated with gravel and sand bed systems respectively, such that competence and capacity limits are exceeded and flow divides around sediment stored in the channel zone.

In contrast to various mid-channel sedimentation features, rivers that are more readily able to accommodate their sediment load or have lower available energy are commonly characterized by bank-attached bars. Dependent on channel/flow alignment, lateral and POINT BARS are found at channel margins under both sand and gravel conditions. These features record sediment accretion on the convex slopes of river bends. Lateral bars tend to occur along straighter river reaches, while point bars are formed on bends. Scroll bars on the inside of bends may form a distinct element in themselves, while former positions of the channel may be recorded by a series of accretionary ridges and intervening swales (Nanson 1980). A range of bar forms have also been characterized for laterally constrained sinuous channels, such as point dunes (Hickin 1969), gravel counterpoint bars (Smith 1987) and convex bar deposits (Goodwin and Steidtmann 1981).

Most river bars are not simple unit features, but are complex, compound features made up of a mosaic of depositional forms such as bar platforms, ridges, chute channels, etc. Compound bars can be differentiated into mid-channel and bank-attached forms. On mid-channel compound bars, chute channels may dissect the bar surface into a chaotic pattern of remnant units. Variants of within-channel compound bar features in sand-bed channels include linguoid bars (Collinson 1970), macroforms (Crowley 1983), sand flats (Cant and Walker 1978), sand waves (Coleman 1969) and sandsheets (Smith 1970). Vegetated mid-channel compound bars are referred to as islands. The array of smaller-scale geomorphic units that make up an island provides key insights into its formation and reworking. On bank-attached compound bars a range of erosional and depositional features such as chute channels, ridges and ramps can be formed under varying flow conditions. Chute channels short-circuit the main body of flow in a river. Enlargement of the chute channel and plugging of the old channel proceed gradually, resulting in a chute cut-off. Because of the small angular difference between the old channel and the chute channel, the stream continues to flow through the old channel for some time, depositing bedload sediment at the upstream and downstream ends and on the floor and sides until terminal closure of the cut-off is complete. Ramp units are depositional forms that result from deposition of coarse gravels within a partially-filled chute channel. These features have a steep upstream facing surface that effectively plugs the chute channel, disconnecting it from the downstream outlet. These chute channel fills are notably straighter in outline than either meander cut-offs or swales.

In quite different environmental settings, bedrock accretionary forms occur on low slopes. These bedrock core bars are characterized by bedrock ridges atop which alluvial materials have been deposited during the waning stages of floods. A positive feedback mechanism is induced when vegetation colonizes these surfaces inducing further deposition and the vertical building of a bar feature. These features are common along bedrock-anastomosing rivers (e.g. Van Niekerk *et al.* 1999).

References

Brierley, G.J. (1991) Bar sedimentology of the Squamish River, British Columbia: definition and application of morphostratigraphic units, *Journal of Sedimentary Petrology* 61, 211–225.

——(1996) Channel morphology and element assemblages: A constructivist approach to facies modelling, in P. Carling, and M. Dawson (eds) *Advances in Fluvial Dynamics and Stratigraphy*, 263–298, Chichester: Wiley Interscience.

Cant, D.J. and Walker, R.G. (1978) Fluvial processes and facies sequences in the sandy braided South Saskatchewan River, Canada, *Sedimentology* 25, 625–648.

Church, M. and Jones, D. (1982) Channel bars in gravel-bed rivers, in R.D. Hey, J.C. Bathurst and C.R. Thorne (eds) *Gravel-bed Rivers: Fluvial Processes, Engineering and Management*, 291–338, Chichester: Wiley.

Coleman, J.D. (1969) Brahmaputra River: channel processes and sedimentation, *Sedimentary Geology* 3, 129–239.

Collinson, J.D. (1970) Bedforms of the Tana River: Norway, *Geografiska Annaler* 52A, 31–55.

Crowley, K.D. (1983) Large-scale bed configurations (macroforms), Platte River Basin, Colorado and Nebraska: primary structures and formative processes, *Geological Society of America Bulletin* 94, 117–133.

Goodwin, C.G. and Steidtmann, J.R. (1981) The convex bar: member of the alluvial channel side-bar continuum, *Journal of Sedimentary Petrology* 51, 129–136.

Hickin, E.J. (1969) A newly identified process of point bar formation in natural streams, *American Journal of Science* 267, 999–1,010.

Nanson, G.C. (1980) Point bar and floodplain formation of the meandering Beatton River, northeastern British Columbia, Canada, *Sedimentology* 27, 3–30.

Smith, N.D. (1970) The braided stream depositional environment: comparison of the Platte River with some Silurian clastic rocks, north central Appalachians, *Geological Society of America Bulletin* 81, 2,993–3,014.

Smith, S.A. (1987) Gravel counterpoint bars: examples from the River Tywi, South Wales, in F.G. Ethridge, R.M. Flores and M.D. Harvey (eds) *Recent Developments in Fluvial Sedimentology*, Society of Economic Paleontologists and Mineralogists Special Publication Number 39, 75–81.

Van Niekerk, A.W., Heritage, G.L., Broadhurst, L.J. and Moon, B.P. (1999) Bedrock anastomosing channel systems: morphology and dynamics in the Sabie River, Mpumalanga Province, South Africa, in A.J. Miller, and A. Gupta (eds) *Varieties of Fluvial Form*, 33–51; Chichester: Wiley.

SEE ALSO: channel, alluvial; fluvial geomorphology, point bar

KIRSTIE FRYIRS AND GARY BRIERLEY

BARCHAN

Barchan is an active crescentic-shaped dune (see DUNE, AEOLIAN), developing in areas of unidirectional winds and limited sand supply. Most barchans are arranged in belts, in which they follow each other. Such belts contain dunes of different sizes: small dunes, which do not exceed several tens of metres in length or width and a few metres in height, develop in a short time; and meso-dunes, which rise up to 40 m and attain several hundred metres in length or width. In various deserts such as in the Western Desert of Egypt (Embabi 1982), Qatar, Peru and California, dimensions of simple barchans show strong linear allometric relationships, such as between length of windward side and dune height, and dune length and width of horns (Plate 10). Mega-barchans that attain heights up to 120 m and lengths of 2–4 km are less common than small and meso-barchans, and are recorded in areas such as the northern parts of Rub' al-Khali in Arabia, and Taklamakan. Sand supply, wind environment, atmospheric motion and age are the main controlling factors of barchan size.

Slope form is concave–convex on the windward side of simple barchans, with angles varying between 1° and 10°. As the barchan grows in size, concave segments occupy a higher percentage of the total length of the windward side. The form of the leeward side changes from convex–concave to convex–straight to mostly straight when it acquires the angle of repose.

Plate 10 Barchans are crescentic dunes, the horns of which point downwind. These examples are in the Western Desert of Egypt in the Kharga Depression

The internal structure of barchans reflects the dynamics of sand removal and deposition on both dune sides. The dominant structure is composed of thin steeply dipping cross-strata resulting from grainflow and grainfall deposited on the slip face, and is preserved as the dune migrates downwind. A secondary horizontal to low dipping structure develops due to deposition on the top of the dune. The sets of cross-strata are separated by horizontal to steeply dipping bounding surfaces.

Barchan moves in the downwind direction due to the dynamics of sand removal and deposition on dune sides. Sand is removed from the lower part of the windward side, and is deposited on the dune crest or on the leeward side. Accumulation of sand on the dune crest leads to periodic sand avalanches on the surface of the slip face. By time, sand removed from the windward side is deposited on the leeward side/slip face, resulting in dune advancement in the downwind direction. Average annual net migration of barchans varies between a few metres to 100 m. Wind energy, dune size and surface relief are the most significant controlling factors in dune advancement. As barchans move forward, they encroach on highways, railways, fields and buildings, representing a permanent hazard to all sorts of human activities, unless checked.

References

Embabi, N.S. (1982) Barchans of the Kharga Depression, in F. El-Baz and T.A. Maxwell (eds) *Desert Landforms of SW Egypt: A Basis for Comparison with Mars*, 141–155, Washington, DC: NASA.

Embabi, N.S. and Ashour, M.M. (1993) Barchan dunes in Qatar, *Journal of Arid Environment* 25, 49–69.

Lancaster, N. (1995) *Geomorphology of Desert Dunes*, London: Routledge.

SEE ALSO: dune, aeolian

NABIL S. EMBABI

BARRIER AND BARRIER ISLAND

Coastal barriers and SPITS are often regarded as similar coastal forms in terms of beach deposition projecting across coastal re-entrants/bays. While barriers tend to bridge the re-entrant by joining the mainland at each end, spits are only attached at one end. This distinction is not rigid, as many barriers show cross-barrier breaks or breaches through which the sea may enter on a permanent or intermittent basis, thus forming barrier segments or barrier islands. The degree of segmentation required to allow the 'island' nomenclature is not defined. Studies on the US eastern seaboard (Johnson 1925; Hoyt 1967; Kraft 1971; Leatherman 1979) have concentrated on barrier formation and reworking, though studies on barrier stability and defending barriers in the face of rising sea levels have come to the fore, pressured by the extensive and expensive real estate located on them (Titus 1990).

Coastal barriers are complex constructional morphology involving deposition by waves, wave-generated currents, tidal currents and wind activity (Hayes 1979). By morphological definition a barrier must exhibit two morpho-dynamic environments/units: a seaward beach face and a landward facing back-barrier slope (Plate 11). These two units are exposed most clearly when the barrier is gravel-dominated (Orford *et al.* 1996). As sand becomes the dominant sediment, a third environment comprising aeolian dunes can appear at the top of the beach face (barrier crest) and spread onto the backslope. Current flows may have been responsible for the initial submarine platform under the barrier, but over time and with sea-level rise, wave action forcing barrier onshore-migration generates much of the later barrier's basal-stratigraphy in combination with fine sedimentation characteristic of the low-energy back-barrier bay. Tidal currents become influential once barrier islands appear.

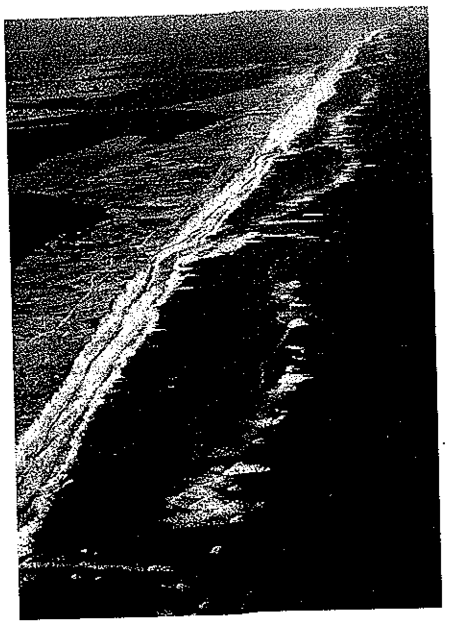

Plate 11 Sand-dominated coastal barrier, Long Island, New York State. Photo by permission of Whittles Publishing, Caithness

Some barriers along the US coast are thought to have evolved out of spits (Fisher 1967). However, other barrier origin mechanisms have been proposed that involve onshore rather than longshore sediment supply. Offshore shoals and longshore bars have been proposed as accretional cores to barriers, but as bars are now recognized as being beach face dependent, it is unlikely that bars can appear before the barrier defines the beach face. Roy *et al.* (1994) suggested that some Australian barriers emerged through shoreface accretion when sea level achieved stationarity after major early-Holocene fluctuations and the shelf area was reworked with excess sediment moving onshore. This latter position sees barriers accreting throughout their length, their plan-view position being set by wave refraction of constructional swell waves. Regardless of origin per se, barriers should be seen as multi-phase morphology relating to changes in controlling variables. Barriers should also be seen in terms of a palimpsest imposed by the structure and roughness of the terrestrial platform that they are superimposed on.

Sediment type and supply are major controls on barrier development with a behavioural distinction to be drawn between sand-dominated barriers (SDB) and gravel-dominated barriers (GDB). This distinction has a spatial basis with GDB being more prevalent in mid-upper latitudes compared to SDB, a dominance reflecting the greater potential of coarse material in high latitudes as a residue of late Quaternary glacigenic processes. SDB were often associated with low angle coastal plains, but barriers are potentially viable wherever sand sinks can be found, e.g. Holocene restructuring of deltas.

Barriers are generally seen as having a Holocene timescale, though this may be truncated to periods since the last major eustatic sea-level variation occurred. In particular many barriers around the North Atlantic have a history commensurate with the mid-Holocene decline in the rate of relative sea-level (RSL) rise. This emphasizes RSL as a datum control on wave activity and barrier development, and the rate of RSL change as a control on the tempo of barrier migration. Early work on the US east coast barriers identified their development with a Holocene TRANSGRESSION that swept up available sediment and concentrated it into the barriers. This perspective has been challenged in that, although it sounds intuitively correct, the actual initial mechanism for onshore concentration of sediment in the surf zone with rising RSL has not been verified, hence the switch to spit elongation as a more coherent model of barrier building in the face of rising RSL. An alternative perspective is to consider aggraded barriers as a consequence of falling RSL. This suggestion has been made for some Florida and Texas barriers, identifying a higher than present sea level during the mid to late Holocene, as a consequence of which barriers developed and aggraded during the subsequent regression. The lack of an obvious mechanism for barrier build-up during a transgression should not be confused with the more understandable behaviour of an existing barrier during subsequent transgressive phases. Jennings *et al.* (1998) suggest that the longshore coherency of GDB relates to the rate of RSL rise: slower rates ($<2\,\mathrm{mm\,a^{-1}}$) mean reduced longshore sediment supply and the cannibalization of existing barrier segments to the point of barrier breaching. Barrier migration rates generally relate to RSL rise rate, though severely reduced GDB may be overwhelmed by the ambient RSL rise, and flattened, to be rebuilt further onshore (i.e. overstepped; Forbes *et al.* 1991).

Wave climate is a major constraint on barriers. Many barriers are prominent in areas that are dominated by oceanic swell. These swells are likely to be transformed as constructional breakers, refracting parallel to existing shorelines and minimizing offshore sediment losses. This does not mean that barriers cannot emerge in storm-wave dominated areas, indeed such areas often show GDB, which tend to move onshore during storms, e.g. Atlantic Canada (Orford *et al.* 1996) and Patagonia (Isla and Bujalesky 2000). Storms are of crucial importance to barrier development. Most storms are expected to work at moving sediment offshore (gravel barriers aside), however as the severity of the storm goes upscale, then the emphasis of sediment transport switches from offshore to onshore. This threshold is reached when run-up physically reaches beyond the barrier crest and transports beach face material over the crest on to the back slope. This process is overwashing and its product is known as washover sedimentation. The latter is most obvious where overwash generates distinct flow channels (throats) through the crest and down the back slope ending in fan splay deposition beyond the previous back-barrier shoreline; such splays are the basis for barrier retreat. The position of throats and fans are partially dictated by the longshore gradients of the breaking wave and morphology of the barrier's seaward face. Beach faces showing cuspate morphology can preferentially set a rhythmic spacing to overwash, which in turn forces a consistent barrier retreat. As storm severity increases then the volume and depth of surface flows over the crest cause lateral extension of throats to the point of coalescence and the overt channels are lost in the face of generalized mobilization (sluicing overwash) of the barrier top into the back-barrier bay area. As sediment is washed into the back-barrier bay, it helps to build up a sub-aqueous sedimentation base for later marsh sedimentation (pads). These pads clearly help to fill in an accommodation space over which the barrier will migrate, such that the shallower the back-barrier area becomes the faster the potential for the barrier to migrate. Some controversy has been generated by the perceived influence of dunes on this migration (and hence survival) process, as dunes will block overwash, or spatially defuse the longshore consistency of overwash and hence reduce migration rate. This led to a short-lived management policy on the US barriers of advocating bulldozing the dune so as to promote overwash – scarcely a recipe for short-term barrier stability. The reverse is now in favour, that of promoting dune sedimentation as a 'sustainable' coastal defence.

Severe storms are also important for the development of cross-barrier breaches. US east coast barriers are vulnerable to hurricanes generating storm surge flow whose elevations are higher than the barrier crest. Hurricane overwash can back-up in the back-barrier bays, impounded by onshore surges, and flow laterally to escape; (1) through old breaches in the barrier; (2) along old overwash channels and (3) at any topographic low point on the barrier that erodes to form a new cross-barrier breach. These breaches can be quickly sealed up by post-hurricane beach face longshore sediment transport, or breaches may persist over decades. Sediment can be transported either way through breaches and surge deposits on the bay side can act as shallow water platforms for future barrier retreat. It is these overwash events that are responsible for most coarser sediment (i.e. high-energy) found in low-energy back-barrier environments.

Tidal range is considered as an indirect control of barrier development. As the tidal range expands, then so does the effectiveness of the tidal current flow regime. Increased tidal prisms maintain the storm-generated breaches. This tidal regime prevents post-storm breach healing and diverts longshore sediment into stores formed as flood or ebb deltas offset from the barrier. The larger the tidal range, the more breaches can be maintained in terms of hydraulic efficiency. Texas has one of the longest barriers in the USA – Padre Island is over 100 km in length and though subjected to hurricane attacks and suffering some breaches (mostly sealed) does not have a sufficient tidal prism (micro-tidal) to maintain the hurricane openings. New Jersey also has storm breaches with a low tidal range, but low longshore sediment availability reinforced by human interference is holding open breaches that would be closed elsewhere. South Carolina experiences a meso-tidal regime that maintains a greater longshore density of breaches or tidal passes, sufficient to define barrier islands. As a concomitant of inlet development and maintenance, there are substantial flood and ebb deltas linking the barrier islands into a hydraulically efficient network conditioned by tidal prism. If the prism alters due to back-barrier reclamation then inlet dimensions will alter, e.g. the Friesian Islands, German Wadden Sea (FitzGerald *et al.* 1984). The ebb/flow characteristics may be conditioned by saltmarsh growth within the bay, acting as a retardant to balanced flood/ebb tidal flow asymmetry. It is rare to find barriers in macro-tidal ranges, but when they do occur (north Norfolk, England) the forcing of the barrier is more to do with longshore sediment supply than tidal inlet forcing, however subsequent evolution in the face of diminishing sediment may mean that the potential for segmentation is great and the apparent morphology of barrier islands may be superimposed.

References

Fisher, J.J. (1967) Origin of barrier island chain shorelines, Middle Atlantic states (abs.) *Geological Society of America*, Special Paper 115, 66–67.

FitzGerald, D.M., Penland, S. and Nummedal, D. (1984) Changes in tidal inlet geometry due to back-barrier filling: East Friesian islands, West Germany, *Shore and Beach*, 52, 3–8.

Forbes, D.L., Taylor, R.B., Orford, J.D., Carter, R.W.G. and Shaw, J. (1991) Gravel barrier migration and overstepping, *Marine Geology* 97, 305–313.

Hayes, M.O. (1979) Barrier island morphology as a function of tidal and wave regime, in S.P. Leatherman (ed.) *Barrier Islands*, New York: Academic Press.

Hoyt, J.H. (1967) Barrier island formation, *Geological Society of America Bulletin* 78, 1,125–1,136.

Isla, F.I. and Bujalesky, G.G. (2000) Cannibalisation of Holocene gravel beach-ridge plains, northern Tierra del Fuego, Argentina, *Marine Geology* 170, 105–122.

Jennings, S.C., Orford, J.D., Canti, M., Devoy, R.J.N. and Straker, V. (1998) The role of relative sea-level rise and changing sediment supply on Holocene gravel barrier development; the example of Porlock, Somerset, UK, *Holocene* 8, 165–181.

Johnson, D.W. (1925) *The New England Arcadian Shoreline*, New York: Wiley.

Kraft, J.C. (1971) Sedimentary facies pattern and geologic history of Holocene marine transgression, *Geological Society of America Bulletin* 82, 2,131–2,158.

Leatherman, S.P. (ed) (1979) *Barrier Islands*, New York: Academic Press.

Orford, J.D., Carter, R.W.G. and Jennings, S.C. (1996) Control domains and morphological phases in gravel-dominated coastal barriers, *Journal of Coastal Research* 12, 589–605.

Roy, P.S., Cowell, M.A., Ferland, M.A. and Thom, B.G. (1994) Wave dominated coasts, in R.W.G. Carter and C.D. Woodroffe (eds) *Coastal Evolution: Late Quaternary Shoreline Morphodynamics*, 121–186, Cambridge: Cambridge University Press.

Titus, J. (1990) Greenhouse effect, sea-level rise and barrier island: case study of Long Beach Island New Jersey, *Coastal Management* 18, 65–90.

JULIAN ORFORD

BASE LEVEL

The concept that there is an effective lower limit to erosional processes. Powell (1875) first named the concept: 'we may consider the level of the sea to be a grand base level, below which the dry lands cannot be eroded; but we may also have, for local and temporary purposes, other base levels of erosion, which are the levels of the beds of the principal streams which carry away the products of erosion.' Chorley and Beckinsale (1968) recognized four main interpretations of the term.

1. Grand base level or 'ultimate base level' which is the plane surface forming the extension of the sea under the lands.
2. Temporary or structural base level, whereby there is a limit to downward erosion of an ephemeral character imposed headwards of a resistant outcrop.
3. Base-levelled surface, which is an ultimate or penultimate topographic surface.
4. Local base level, as for example in areas of interior drainage under an arid cycle.

The first of these usages occupied a fundamental place in the cycle of erosion concept of W.M. Davis (1902). Base-level changes are also crucial in the study of fluvial terraces, deltas and other depositional systems (Koss *et al.* 1994). They can result from tectonic activity, sea-level change and river capture (Mather 2000).

References

Chorley, R.J. and Beckinsale, R.P. (1968) Base-level, in R.W. Fairbridge (ed.) *The Encyclopedia of Geomorphology*, 58–60, New York:Reinhold.

Davis, W.M. (1902) Base-level, grade and peneplain, *Journal of Geology* 10, 77–111.

Koss, J.E., Ethridge, F.G. and Schumm, S.A. (1994) An experimental study of the effect of base-level change on fluvial, coastal plain and shelf systems, *Journal of Sedimentary Research* B, 64, 90–98.

Mather, A.E. (2000) Adjustment of a drainage network to capture induced base-level change. An example from the Sorbas Basin, S.E. Spain, *Geomorphology* 34, 271–289.

Powell, J.W. (1875) *Exploration of the Colorado River of the West*, New York.

A.S. GOUDIE

BEACH

A beach is a wave deposited accumulation of sediment located between modal wave base and the upper swash limit. Beaches may be composed of

fine sand through boulders and may range from low energy, narrow strips of sand lapped by low wind waves, to high energy systems exposed to persistent 2–3 m high swell which breaks across 500 m wide surf zones. Beach systems are also located in all tide ranges, in all latitudes, in all climates and on all manner of coast, from the low coastal plains fronted by long beaches to small pockets of sand wedged in at the base of massive cliffs. Therefore beaches exist in a wide spectrum of wave, tide and sediment combinations and geological settings.

In two dimensions, however, all beaches will contain three dynamic zones – of wave shoaling, wave breaking and swash–backwash. The wave shoaling zone extends from the modal wave base where average waves can entrain and move sediment shoreward, to the outer breakpoint. The wave shoaling zone is dominated by asymmetrical wave orbital motions which produce a concave upward profile sloping less than 1°, dominated by wave ripples which become increasingly three dimensional close to the breaker zone. The depth and width of this zone is dependent on wave height and sediment size. On high energy coasts it extends out to depths of 30 m or more which may lie 2–3 km offshore, while on low energy systems it may only extend to low tide a few metres from the shore. Sediment is often graded across the system and fines seaward.

The surf zone is located between the breakpoint and shoreline. The surf zone has the greatest potential for complex dynamic processes and resulting topography and bedforms. The dominant processes start with wave breaking. Onshore currents are generated by wave bores and orbital currents associated with reformed waves. Longshore currents result from oblique waves and bi-directional rip feeder currents. Offshore flows are driven by wave reflection, discrete rip currents and bed return flow. The width of the surf zone is dependent on surf zone gradient, a function of sand size and wave height. It will be as narrow as a few metres on a steep reflective beach, typically 50–100 m wide on a single bar intermediate beach and up to several hundred metres on a high energy dissipative beach. The presence and number of bars will increase with decreasing wave period and decreasing gradient. Longshore variations in form and processes are driven by longshore changes in wave conditions, and by three-dimensional bar and rip topography.

Surf zone topographic features include shore parallel bars and troughs with waves breaking over the bars and reforming in the troughs. Swell coasts rarely have more than two bars (see BAR, COASTAL), while energetic sea coasts may have several bars. On intermediate beaches with cellular rip circulation, alternating shore transverse bars and rips, also known as crescentic bars, can occur. When present they dominate the inner bar and can lead to a rearranging of the shoreline to produce rhythmic topography. Surf zone BEDFORMS reflect the changing velocity and direction of currents and depth of water, and range from flat bed over the shallow bars, to wave orbital and shore perpendicular current ripples in the troughs, to shore parallel seaward migrating ripples in the rip channels.

The swash zone extends up the beach from the shoreline, from where the wave breaks or bore collapses, to the limit of swash. The swash zone is always an upward sloping zone of wave uprush (swash) and backwash. Its slope is directly related to grain size and inversely related to wave height and may range from 1° to 10°. It is usually featureless, with ripples only produced by strong backwash in the lower swash zone. The swash may, however, be superimposed on high tide beach cusps or a berm, and mesoscale megacusps. Beyond the limit of normal swash is the backshore, a zone of either wind-blown aeolian deposits and/or storm wave overwash.

In three dimensions beaches respond to a greater range of variables and become increasingly complex. First is the alignment of the shoreline to the dominant wave crest, which produces swash aligned beaches. As waves refract around headlands and nearshore topography, the wave crests bend to parallel the contours. At the shoreline the beach also moves to parallel the wave crests so as to minimize longshore transport, and thereby produce a more stable shoreline in equilibrium with the wave crest. Where waves arrive at a persistent oblique angle to the shore, particularly on long beaches, then sediment is moved downdrift by the longshore surf zone currents generated by the waves, producing a drift aligned shore.

Beach type

Beach type refers to the morphodynamic character of a beach system, which is a product of the interaction of waves, tide and sediment. Beaches may be of three types: wave dominated, tide modified and tide dominated. Wave-dominated beaches occur where waves are high relative to the tide range. This can be defined quantitatively by the relative tide range

$$RTR = TR/H_b \qquad (1)$$

where TR is the spring tide range and H_b the average breaker wave height. When $RTR < 3$ beaches are tide dominated, when $3 < RTR < 15$ they are tide modified and when the $RTR > 15$ they become tide dominated.

Within each of these beach types a range of wave and sediment combinations can occur which will influence the actual state of the beach. The dimensionless fall velocity

$$\Omega = H_b/T\ W_s \qquad (2)$$

where T is wave period (s) and Ws sediment fall velocity ($m\,s^{-1}$) can be used to quantify beach state. They range from the lower energy reflective ($\Omega < 1$) favoured by low waves, longer periods and coarser sediments, to dissipative ($\Omega > 6$) with high waves, shorter periods and fine sand. In between are the more rip dominated intermediate beaches ($\Omega = 2$–5) produced by moderate to high wave conditions.

WAVE-DOMINATED BEACHES

Wave-dominated beaches consist of three types: reflective, intermediate and dissipative.

Reflective beaches

Reflective beaches are produced by combinations of lower waves, longer periods and particularly coarser sands. They occur on sandy open swell coasts when waves average less than 0.5 m, and on all coasts when beach sediments are composed of coarse sand or coarser, including all gravel and boulder beaches, even under higher waves. However, they are all characterized by a concave upward nearshore zone of wave shoaling that extends to the shoreline. Waves then break by plunging and/or surging across the base of the beach face. The ensuing strong swash rushes up the beach, combining with the coarse sediment to build a steep beach face (4°–10°), commonly capped by well-developed beach cusps and/or a *berm* (Plate 12), possibly backed by a runnel where the high tide swash may temporarily accumulate. When the sediment consists of a range of grain sizes, the coarser grains accumulate as a coarser steep *step* below the zone of wave breaking, at the base of the beach face.

The cusps are a product of cellular circulation on the high tide beach resulting from sub-harmonic

Plate 12 A lower energy reflective beach with wave surging up the moderately steep beach, Horseshoe Bay, South Australia (Andrew D. Short)

edge waves produced from the interaction of the incoming and reflected backwash. The high degree of incident wave reflection off the beach face is responsible for the naming of this beach type, i.e. reflective. Apart from the cosmetic beach cusps and swash circulation these are essentially two-dimensional beaches with no longshore variation in either processes or morphology. On sand beaches they also represent the lower energy end of the beach spectrum and as such are relatively stable systems only responding to an increase in wave height. Such an increase induces a growth in the swash energy and erosion of the swash zone.

Intermediate beaches

Intermediate beaches are called such as they represent a suite of beach types between the lower energy reflective and higher energy dissipative. They are the beaches that form under moderate to high waves, on swell and sea coasts, in fine to medium sand. The two most distinguishing characteristics of intermediate beaches are (1) a surf zone, and (2) cellular rip circulation (see RIP CURRENT) commonly associated with rhythmic bar and beach topography (Plate 13). Since intermediate beaches can occur across a wide range of wave conditions, they consist of four beach states ranging from the lower energy low tide terrace to the rip dominated transverse bar and rip and rhythmic bar and beach, and the high energy straighter longshore bar and trough.

Intermediate beaches are controlled by processes related to wave dissipation across the surf zone which transfers energy from incident

Plate 13 Well-developed transverse bar and rips, Lighthouse Beach, New South Wales, Australia (Andrew D. Short)

waves with periods of 2 to 20 s, to longer infragravity waves with periods >30 s. As a consequence, incoming long waves associated with wave groupiness, increase in energy and amplitude across the surf zone and are manifest at the shoreline as wave set up (crest) and set down (trough). The long wave then reflects off the beach leading an interaction between the incoming and outgoing waves to produce a standing wave across the surf zone. It is believed that standing edge waves trapped in the surf zone are responsible for the cellular circulation that develops into rip current circulation and associated transverse bars and rips. These are in turn responsible for the high degree of spatial and temporal variability in intermediate beach morphodynamics.

The low tide terrace (or ridge and runnel) beaches are characterized by a continuous attached bar or terrace located at low tide. They form under lower waves (0.5–1 m) and usually undergo temporal variation between low tide when the waves break and dissipate across the bar, while at high tide they may remain unbroken and surge up the reflective high tide beach face. Weak rips may occur at mid to low tide.

Transverse bar and rip beaches form under moderate waves (1–1.5 m) on swell coasts and consist of well-developed rip channels, which are separated by shallow bars, the bars attached and perpendicular or transverse to the beach (Plate 13). Variable wave breaking and refraction across the shallow bars and deeper rip channels lead to a longshore variation in swash height and approach, which reworks the beach to form prominent megacusp horns in lee of bars, and embayments in lee of channels. Water tends to flow shoreward over the bars, then into the rip feeder channels. The flow moves close to the shoreline and converges laterally in the rip embayment. It then moves seaward in the rip channel as a relatively narrow (few metres), strong flow (0.5–$1\,m\,s^{-1}$), called a *rip current*. This beach state has extreme spatial-longshore variation in wave breaking, surf zone and swash circulation and beach and surf zone topography, leading to a highly unstable and variable beach system.

The rhythmic bar and beach state forms during periods of moderate to high waves (1.5–2 m) on swell coasts. The high waves lead to greater surf zone discharges that require deeper and wider rip feeder and rip channels to accommodate the flows. Rips flow in well-developed rip channels, separated by transverse bars, however the bars are detached from the beach by the wider feeder channels.

The longshore bar and trough systems are a product of periods of higher waves which excavate a continuous longshore trough between the bar and the beach. Waves break heavily on the outer bar, reform in the trough and then break again at the shoreline, often producing a steep reflective beach face (coarser sand) or low tide terrace (finer sand). Surf zone circulation consists of both cellular rip flows as well as increasingly shore normal bed return flows (see below).

Dissipative beaches

Dissipative beaches represent the high-energy end of the beach spectrum. They occur in areas of high waves, prefer short wave periods, and must have fine sand. They are relatively common in exposed sea environments where occasional periods of high, short storm waves produce multi-barred dissipative beach systems, as in the North and Baltic seas. They also occur on high-energy mid-latitude swell and storm wave coasts as in northwest USA, southern Africa, southern Australia and New Zealand. On swell coasts waves must exceed 2–3 m for weeks to generate fully dissipative beaches. They are characterized by a wide long gradient beach face and surf zone, with two and more shore parallel bars forming across the surf zone (Plate 14). The

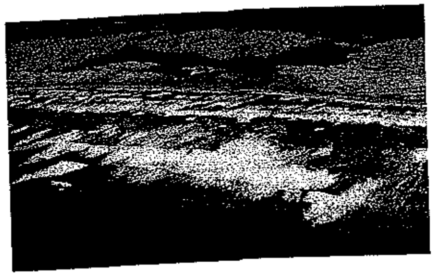

Plate 14 High energy dissipative beach containing an inner bar, trough and wide outer bar, Dog Fence Beach, South Australia (Andrew D. Short)

low gradient is a product of both the fine sand, as well as the dominance of lower frequency infragravity swash and surf zone circulation, which act to plane down the beach. The name comes from the fact that waves dissipate their energy across the many bars and wide surf zone.

The dissipation of the incident wave energy leads to a growth in the longer period infragravity energy, which becomes manifest as a strong set up and set down at the shoreline. The standing wave generated by the interaction of the incoming and outgoing waves may have two and more nodes across the surf zone. It is believed that the bar crests form under the standing wave nodes and troughs under the antinodes. Surf zone circulation is vertically segregated. Wave bores move water shoreward toward the surface of the water column. This water builds up against the shoreline as wave set up. As it sets down the return flows tends to concentrate toward the bed, which propels a current across the bed of the surf zone (below the wave bores) called *bed return flow*.

Like the reflective end of the beach spectrum dissipative beaches are remarkably stable systems. They are designed to accommodate high waves, and can accommodate still higher waves by simply widening their surf zone and increasing the amplitude of the standing waves, while periods of lower waves are often too short to permit substantial onshore sediment migration.

TIDE-MODIFIED BEACHES

Most of the world's beaches are affected by tide. On most open coasts where tides are low (<2 m) waves dominate and tidal impacts are minimized. However, as tide range increases and/or wave height decreases tidal influences become increasingly important. To accommodate these influences beaches, still by definition wave-formed, can be divided into tide-modified and tide-dominated types, as defined by equation 1 (see p. 64).

The major impact of increasing tide range is to shift the location of the shoreline between high and low tide, which – depending on the shoreline gradient – will be tens to hundreds of metres. This shift not only moves the shoreline and accompanying swash zone, but also the surf zone, if present, and the nearshore zone. While wave-dominated beaches have a relatively 'fixed' swash–surf–nearshore zone, on tide-modified beaches they are more mobile and transient. The net result is a smearing of the three dominant wave processes of shoaling (nearshore zone), breaking (surf zone) and swash (swash zone). A section of intertidal beach can be exposed to all three processes at different states of the tide. Second, because all three zones are mobile, except for a brief period at high and low tide, there is a reduction in the time any one process can fully imprint its dynamics on a particular part of the beach. As a consequence there is a tendency for swash processes only to dominate the spring high tide beach, for surf zone processes only to dominate the beach morphology around low tide, during the turn of the low tide, while shoaling wave processes become increasingly dominant overall, producing a lower gradient, featureless, concave beach cross section.

Tide-modified beaches can contain three states. When waves are lower ($\Omega < 1$–2) they consist of a steep reflective high tide beach face fronted by a wide low gradient low tide terrace, often with a sharp break in slope between the two. At low tide waves dissipate across the terrace, while at high tide they pass unbroken across the now submerged terrace to surge up the steep reflective high tide beach. In areas of moderate waves ($\Omega = 2$–5) and tide range (RTR = 3–7) the tide-modified beaches consist of a high tide reflective beach, a usually wider intertidal zone, and a low tide zone dominated by surf zone morphology, which may include transverse and rhythmic bars and rips. In moderate energy sea environments a series of shore parallel ridges and runnels may develop (Plate 15). Higher energy tide-modified beaches ($\Omega > 5$) composed of fine sand are characterized by a wide, low gradient concave upward, flat and featureless, beach and intertidal system, called an ultradissipative beach (Plate 16).

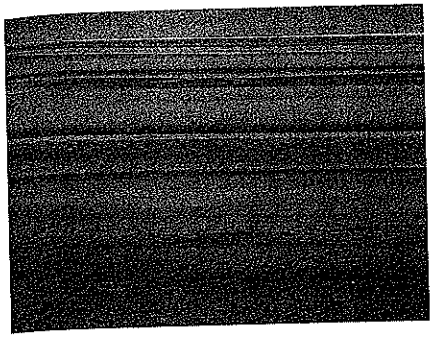

Plate 15 Reflective high tide beach (foreground) fronted by three ridges and runnels, Omaha Beach, France (Andrew D. Short)

Plate 16 Rhossili Beach, Wales, a high energy ultradissipative tide-modified beach, shown here at low tide (Andrew D. Short)

TIDE-DOMINATED BEACHES

Tide-dominated beaches occur when the RTR > 15, that is the tide range is more than 15 times the wave height. As the maximum global tide range is about 12 m, and usually much less, this means that most tide-dominated beaches also receive low (<1 m) to very low waves, and are commonly dominated by locally generated wind waves. Tide-dominated beaches are characterized by a usually steep, narrow high tide beach, and a wide, low gradient (<1°) sandy intertidal zone, which in temperate to tropical locations is usually bordered by subtidal seagrass meadows. They consist with decreasing wave energy of three states: (1) a beach and sand ridges, containing multiple, low amplitude, shore parallel sand ridges across the intertidal; (2) a reflective beach fronted by a usually wide, flat, featureless intertidal sand flat; (3) a tidal sand flat, which is a beach in so far as it has a high tide beach. However, the fronting often wide tidal flats are dominated by the tides and not waves, and may contain intertidal biota and tidal drainage features. It is part of the transition between the beaches and the often muddy, tidal flats.

Further reading

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

Hardisty, J. (1990) *Beaches: Form and Process*, London: Unwin and Hyman.

Horikawa, K. (ed.) (1988) *Nearshore Dynamics and Coastal Processes*, Tokyo: University of Tokyo Press.

Komar, P.D. (1998) *Beach Processes and Sedimentation*, 2nd edition, Upper Saddle River, NJ: Prentice Hall.

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

SEE ALSO: bar, coastal; bedform; rip current

ANDREW D. SHORT

BEACH CUSP

Beach cusps are crescentic concave-seaward regularly spaced features occurring along the shoreline. The term has been used for features with spacing ranging from 10 cm to many hundreds of metres, although larger examples tend to be called rhythmic beach features with the term swash cusp being used for features with a spacing less than tens of metres (Hughes and Turner 1999). Beach cusps (swash cusps) are most commonly associated with medium to coarse sands, shingle or mixed sand-shingle sediments, on steep beaches demonstrating significant wave reflection. Their amplitude ranges from almost zero to over 1 m. On beaches with high tidal range, multiple sets of cusps may be present at different levels. Beach cusps consist of embayments or swales separated by triangular horns which are normally comprised of coarser sediments.

A number of different swash flow patterns in and around cusps have been reported in the literature. Under low energy conditions, oscillatory flows (with swash largely unaffected by cusp morphology), horn divergent flows (uprush flow separation at the horn with water returning from the embayment), and horn convergent flows (uprush entering the cusp in the embayment and returning along the sides of the horn, converging at its apex) have all been reported. Under high energy conditions sweeping flow (alongshore directed water movement) and swash-jet flows (where incoming swash is held back by the backwash until it develops sufficient head to break through as a jet in the centre of the embayment) can also occur (Masselink and Pattiaratchi 1998).

Numerous conflicting ideas have been proposed for the mode of formation of beach cusps, including processes of accretion, processes of erosion or a combination of both, and theories based on instabilities on wave breaking, alongshore sediment transport, and intersecting wave trains. Debate continues on the formation of beach cusps, with two theories based on fundamentally different mechanisms dominating.

Cusp development caused by standing edge WAVES at either twice the period (subharmonic) of, or synchronous with the incident waves has been proposed. This hypothesis can explain regular spacing (cusp spacing equal to edge wave length for synchronous edge waves, or half edge wave length for sub-harmonic edge waves), but can also explain complex quasi-regular spacing if more than one mode of edge wave is present. Werner and Fink (1993), however, proposed a self-organization model for beach cusp formation, with topographic depressions resulting in feedback mechanisms between swash and morphology resulting in self-emergence of beach cusps. Cusp spacing is proposed as being proportional to swash excursion length. Predications of similar beach cusp spacing based on the quite distinct self-organization and edge wave theories of cusp formation have meant that field studies have been unable to discriminate between the two models. It is also possible that edge waves may initiate cusp development with self-organization then allowing for the growth of the features. Many field studies have reinforced the importance of feedback processes between swash and morphology (Masselink *et al.* 1997).

References

Hughes, M. and Turner, I. (1999) The beachface, in A.D. Short (ed.) *Handbook of Beach and Shoreface Morphodynamics*, 119–144, Chichester: Wiley.

Masselink, G. and Pattiaratchi, C.B. (1998) Morphological evolution of beach cusps and associated swash circulation patterns, *Marine Geology* 98, 93–113.

Masselink, G., Hegge, B.J. and Pattiaratchi, C.B. (1997) Beach cusp morphodynamics, *Earth Surface Processes and Landforms* 22, 1,139–1,155.

Werner, B.T and Fink, T.M. (1993) Beach cusps as self-organized patterns, *Science* 260, 968–971.

Further reading

Komar, P.D. (1998) *Beach Processes and Sedimentation*, 2nd edition, Upper Saddle River, NJ: Prentice Hall.

SEE ALSO: beach; wave

KEVIN PARNELL

BEACH–DUNE INTERACTION

The original generation of the wave–beach–dune model of beach and dune interactions was formulated by Hesp (1982). It was followed by the publication of a robust micro-tidal beach model with reasonably high predictability (Wright and Short 1984). The beach model enabled scientists to classify micro-tidal beaches into six states with characteristic morphologies, mobility, and modes of erosion and accretion. Subsequent research has extended the original model to meso- and macro-tidal beaches (see BEACH). An understanding of beach and backshore morphology for different surfzone-beach types allowed Hesp (1982) to develop actual and theoretical links between backshore morphology, potential aeolian transport, foredune state and morphology, and dunefield type and development.

Surfzone-beach state

The micro-tidal beach models classified beaches into six states, with the dissipative state at the high wave energy (>2.5 m) extreme and reflective state at the low wave energy (<1 m) extreme. Four intermediate beach states occur between these states. Dissipative beaches are characteristically high wave energy beaches and have the highest potential onshore sediment supply (Hesp 1988). Note, however, that beaches may also be dissipative because of the presence of very fine sand (hence low gradient) or abundant sand, so some dissipative beaches may, in fact, be low wave energy beaches. They are typically wide, display flat to concave morphologies (no berms), low gradients and minimal backshore mobility. The latter refers to the coefficient of variation of mean shoreline position, and in reality refers to the amount of volumetric and profile change the beach and backshore experiences over time and through erosion to accretion phases. Reflective beaches are characteristically low wave energy beaches with low potential onshore wave driven sediment transport. Note that they may also be moderate to high wave energy where sediments are coarse sand to boulders. They are relatively steep, narrow, linear to terraced (i.e. display a berm form) morphologies, with low backshore mobility. Intermediate beaches range from wide, relatively flat beaches with low mobility at the higher energy end of the spectrum, through moderate width beaches with pronounced berms and high mobility to narrow beaches and moderate to low mobility berms at the reflective end of the range. Rips dominate surfzone processes in the intermediate range.

Beach-backshore width and morphology, fetch and potential aeolian transport

Beach width is important in determining fetch which is critical for determining the volume of sand delivered across the backshore and to dunes (Davidson-Arnott 1988). Beach morphology is important because the greater the morphological variability, the more likely that wind velocity decelerations and variations take place across the backshore. Hesp (1982, 1999) showed that the wind flow across a wide, low gradient, dissipative beach displayed minimal flow variation and gradually accelerated across the backshore, thus maximizing potential aeolian transport. The wind flow over the berm crest of an intermediate beach was accelerated but decelerated leeward of the berm crest. High narrow berms typical of some reflective beaches display significant flow disturbance and deceleration leeward of the berm crest (Short and Hesp 1982). Sherman and Lyons (1994) modelled wind flow and potential sediment transport on a flat beach, low berm and high berm profiles, and found that sand transport off the dissipative beach was 20 per cent higher than off the reflective beach if just slope and grain size were taken into account. When moisture content was added, transport rates were nearly two orders of magnitude higher off the dissipative beach compared to the reflective beach. Note, however, that each beach had the same width (100 m wide), whereas actual reflective beaches and many intermediate beaches are considerably narrower.

Beach mobility is important because the greater the beach mobility, the greater the morphological variability. The latter affects the fetch such that the beach width is at times quite narrow, at times quite wide, particularly for intermediate beaches. It is also important because alternating episodes of cut-and-fill result in varying beach morphologies which then affect airflow and sediment transport as indicated above.

Thus, the link between surfzone beach state, aeolian sediment transport and landward dunes is that modal dissipative beaches have maximum potential aeolian sediment transport, reflective beaches minimal potential aeolian sediment transport, and intermediate beaches range from relatively high potential at the dissipative end to low potential at the reflective end. Note that a minimal sediment supply ('minimal' is currently undetermined) is required.

Aeolian sediment transport and foredune morphology

An examination of foredune heights and volumes on dissipative to reflective beaches allows one to examine the validity of the links above. Since established foredunes occupy a foremost backshore position, they are a medium-term indicator of beach and backshore processes. Hesp (1988) measured incipient and established foredune volumetric changes over several years at Myall Lakes National Park in New South Wales, Australia to find that a modal reflective beach with the same wind exposure as a modal dissipative beach received 60 per cent less sand than the dissipative beach over the same survey period. Intermediate beach volumes ranged from relatively high to relatively low between the dissipative and reflective beaches.

Surveys of established foredunes, which have been present for potentially several hundred years, provide further evidence that there is a strong link between surfzone-beach type and foredune height and volume. Hesp (1982, 1988) demonstrated that in the Myall Lakes National Park the smallest established foredunes, with lowest sediment volumes were found on reflective beaches, while the highest and largest foredunes occurred on dissipative beaches. Similar results are reported by Davidson-Arnott and Law (1990). Intermediate beaches followed a trend from low to high volumes on lowest to highest energy intermediate beaches respectively (see reviews in Sherman and Bauer 1993 and Bauer and Sherman 1999).

Foredune ecology

The vegetation cover, species richness and zonation of foredunes is determined by several factors, but sediment supply and sand deposition rate, and salt spray aerosol levels are two important factors (Hesp 1991). Simultaneous studies carried out on adjoining reflective, intermediate and dissipative beaches show that salt spray aerosol levels are related to surfzone-beach type. Dissipative beaches have the widest surfzones, the greatest number of breaking waves, and highest wave heights and the highest salt aerosol levels. Reflective beaches often have only one breaking wave, narrow to very narrow surfzones, and low wave heights and the lowest salt aerosol levels. All other factors being equal, foredune species richness and zonation tends to be greatest and narrowest respectively on reflective beaches (low sediment supply and salt aerosols), and lowest and widest on dissipative beaches (highest sediment supply, high salt aerosol levels) (Hesp 1988).

Foredune stability and type, erosion processes and dunefield development

Foredunes bear a morphological imprint dictated, in part, by modal surfzone-beach erosion and accretion modes, and the wind often accentuates this morphological imprint. Dissipative beaches are typically eroded by swash bores and undertow commonly associated with elevated water levels and storm surge. Beach erosion and dune scarping is laterally continuous alongshore, and at times catastrophic. Hesp (1988) and Short and Hesp (1982) theorized that such laterally continuous alongshore, large-scale foredune scarping would on occasions lead to large-scale foredune destabilization. Transgressive dunefields would most likely result from the breakdown of the large established foredune. In fact, transgressive dunefields are most commonly found on high energy dissipative surfzone-beach systems (e.g. Australian and South African coasts below the tropics; west coast USA; west coast North Island, New Zealand).

Intermediate beaches are characterized by localized, arcuate rip embayment erosion during storms. Such arcuate erosion extends well into the foredune during extreme events resulting in large-scale but localized foredune scarping. Topographic funnelling of the wind may result in the evolution of blowouts and eventually parabolic dunes at these locations. On average, many higher energy intermediate beaches display parabolic dune complexes (Hesp 1982, 1988; Short and Hesp 1982).

On south-east Australian beach systems where overwash events are minor to absent, where sediment supply is generally not limited, and where an aggressive pioneer grass (*Spinifex* sp.) exists, relict foredune plains are common, particularly on the moderate energy intermediate beaches. Here established foredune stability is maintained to various degrees, and progradation over the last 6,000–7,000 years has led to the development of foredune plains.

Reflective beaches are characterized by accentuated swash during storms and laterally continuous alongshore beach erosion. Recovery is fairly rapid. Foredunes remain relatively stable over time, and because they are typically small, with limited sediment supply, little dune transgression results. Thus reflective beaches are characterized by a single foredune, or a few relict foredunes.

Role of sediment supply and other factors

There is no doubt that sediment supply, wind energy, sea-level state (transgressive, stable, regressive), return interval and magnitude of extreme storm events, and Pleistocene inheritance factors will all, at times, and in some places, be a controlling variable in beach–dune interactions. If sediment supply is limited, sea level is rising, and coastal erosion is the general rule, the model as outlined above may not work in part or perhaps at all (Psuty 1988).

References

Bauer, B.O. and Sherman, D.J. (1999) Coastal dune dynamics: problems and prospects, in A.S. Goudie, I. Livingstone and S. Stokes (eds) *Aeolian Environments, Sediments and Landforms*, 71–104, Chichester: Wiley.

Davidson-Arnott, R.G.D. (1988) Temporal and spatial controls on beach/dune interaction, Long Point, Lake Erie, in N.P. Psuty (ed.) *Dune/Beach Interaction, Journal of Coastal Research Special Issue* 3, 131–136.

Davidson-Arnott, R.G.D. and Law, M.N. (1990) Seasonal patterns and controls on sediment supply to coastal foredunes, Long Point, Lake Erie, in K.F. Nordstrom, N.P. Psuty and R.W.G. Carter (eds) *Coastal Dunes: Form and Process*, 177–200, Chichester: Wiley.

Hesp, P.A. (1982) *Morphology and Dynamics of Foredunes in S.E. Australia*, unpublished Ph.D. Thesis, Dept. Geography, University of Sydney.

——(1988) Surfzone, beach and foredune interactions on the Australian south east coast, *Journal of Coastal Research Special Issue* 3, 15–25.

——(1991) Ecological processes and plant adaptations on coastal dunes, *Journal of Arid Environments* 21, 165–191.

——(1999) The beach backshore and beyond, in A.D. Short (ed.) *Handbook of Beach and Shoreface Morphodynamics*, 145–170, Chichester: Wiley.

Psuty, N.P. (1988) Sediment budget and beach/dune interaction, in N.P. Psuty (ed.) Dune/Beach Interaction, *Journal of Coastal Research Special Issue* 3, 1–4.

Sherman, D.J. and Bauer, B.O. (1993) Dynamics of beach–dune interaction, *Progress in Physical Geography* 17, 413–447.

Sherman, D.J. and Lyons, W. (1994) Beach-state controls on aeolian sand delivery to coastal dunes, *Physical Geography* 15, 381–395.

Short, A.D. and Hesp, P.A. (1982) Wave, beach and dune interactions in south eastern Australia, *Marine Geology* 48, 259–284.

Wright, L.D. and Short, A.D. (1984) Morphodynamic variability of beaches and surfzones: A synthesis, *Marine Geology* 56, 92–118.

PATRICK HESP

BEACH NOURISHMENT

Beach nourishment is the act of placing sediment (termed fill) on a beach by artificial means using sources outside the nourished area. Nourishment is primarily conducted to overcome a sediment deficit and create a beach of sufficient width to protect existing buildings and infrastructure from wave attack, but it also can enhance the value of urban locations for tourism or create new natural environments.

Debates have occurred over the cost effectiveness of nourishment and whether nourished beaches erode more rapidly than predicted in design studies (Houston 1991; Pilkey 1992), but projects are increasingly implemented, and nourishment is now the principal option for shore protection in some countries. Nourishment is conducted at all levels of management, from the national level to private homeowner groups. Borrow areas for fill materials include offshore, inlets, backbays, rivers and glaciated uplands. Opportunistic sources from dredging of harbours, marinas, lagoons and inland construction sites are also used (Nordstrom 2000).

Nourishment occurs on the upper beach, the nearshore, offshore on stable berms (designed to alter wave conditions) and active berms (that change shape or migrate onshore), and on existing dunes or on the backbeach to create new dunes. Large-scale nourishment operations on the upper beach commonly use a pipeline to transport a sand/water slurry. Small-scale operations transport sediment in dump trucks. The nourished beach is then often reshaped by bulldozers. The result of upper beach nourishment is a high, wide beach with an unstable shape, but the fill is easy to place, provides good initial protection against wave overwash, and creates a wide recreation platform. Conspicuous losses may subsequently occur on the upper beach as the fill adjusts to a more natural equilibrium shape. Fill sediments on the foreshore are reworked by waves and often become similar to pre-nourished sediments in size and sorting, but fill sediments on the backbeach above the zone of wave reworking retain characteristics that differ from native sediments.

Nearshore nourishment occurs by spraying sediment as a sand/water slurry or dumping it from shallow-draught barges. By placing sediments directly in the dynamic surf zones, losses through time are not visible and aesthetics are not spoiled by different sediments on the backbeach, but a beach nourished this way evolves slowly. Offshore berms are often implemented as disposal areas for sediment dredged from navigation channels, and more study is required to evaluate their use as protection structures.

Sediment bypassing (artificially transporting sediment to the downdrift side of obstacles to littoral drift) and backpassing (artificially transporting sediment from downdrift deposits to updrift eroding zones) may also be considered nourishment projects. Bypassing is gradually becoming more common at inlets where jetties or dredging of navigation channels interrupt longshore sediment transport. Backpassing is now most frequently conducted in small-scale trucking operations, but it may become more significant in the future as ready supplies of external sediment for nourishment projects become exhausted.

Nourishment of the upper beach can alter aeolian transport by (1) increasing the source width for entrainment of sediments; (2) adding fine sediment that is more readily transported; (3) changing moisture-retention characteristics; (4) changing the shape of the beach or dune profile; and (5) changing the likelihood of marine erosion of the incipient foredune (van der Wal 1998). Rapid dune growth can occur on nourished beaches, especially when sand fences and vegetation plantings are used to trap sand.

Dunes may be created directly by mechanically depositing sediment. Most dune nourishment operations place the new fill in front of the existing dune to create a sacrificial structure or on top of it to increase the level of protection against wave overwash; more rarely, a new dune may be built behind an existing foredune (Nordstrom 2000). Dunes built and used as protection structures can evolve into a condition that functions naturally or appears natural in terms of surface vegetation (Nordstrom *et al.* 2002).

Nourished beaches benefit threatened species by providing habitat that would otherwise be unavailable, but detrimental ecological impacts can occur due to (1) mechanical removal of habitat in borrow areas; (2) burial of habitat in nourished areas; (3) increased turbidity and sedimentation; (4) disruptions to foraging, nesting, nursing and breeding; (5) change in sediment characteristics, wave action and beach state; and (6) change in community structure and evolutionary trajectories, including enhancement of undesirable species (Nelson 1993; National Research Council 1995). Detrimental effects are often considered temporary, but little is known about long-term, cumulative impacts and critical thresholds.

Human activities, such as driving on the beach or raking the beach to remove litter, can eliminate incipient topography and vegetation and prevent formation of natural landforms, so true restoration of landforms and habitats may not occur in the absence of controls on subsequent human activities (Nordstrom 2000). The great importance of nourishment as a form of shore protection and as a sediment resource that can evolve into naturally functioning landforms makes this option an important area of future geomorphological research. To be effective, the nourished beach must be considered as a landform in its own right and as a source of sediment for evolution of other landforms landward and downdrift of it, rather than merely as an engineering structure or recreation platform.

References

Houston, J.R. (1991) Beachfill performance, *Shore and Beach* 59(3), 15–24.

National Research Council (1995) *Beach Nourishment and Protection*, Washington, DC: National Academy Press.

Nelson, W.G. (1993) Beach restoration in the southeastern US: environmental effects and biological monitoring, *Ocean and Coastal Management* 19, 157–182.

Nordstrom, K.F. (2000) *Beaches and Dunes of Developed Coasts*, Cambridge: Cambridge University Press.

Nordstrom, K.F., Jackson, N.L., Bruno, M.S. and de Butts, H.A. (2002) Municipal initiatives for managing dunes in coastal residential areas: a case study of Avalon, New Jersey, USA, *Geomorphology*, 48, 147–162.

Pilkey, O.H. (1992) Another view of beachfill performance, *Shore and Beach* 60(2), 20–25.

van der Wal, D. (1998) The impact of the grain-size distribution of nourishment sand on aeolian sand transport, *Journal of Coastal Research* 14, 620–631.

KARL F. NORDSTROM

BEACH RIDGE

Beach ridges are azonal accumulation forms created on the shores of seas and lakes. They are usually subparallel ridges of sand, gravel or pebble, as well as detritus of shell, situated in the foreshore zone, which is the boundary of low and high water's range. Older, subfossil complexes of beach ridges may actually appear in the backshore zone, which lies above the high water's range. Beach ridges forming at the present day are roughly parallel to the coast. The height of their crests is usually a little bit higher than mean high tide or storm, and the bottom of the adjacent troughs or swales have elevations not far from mean low water (Stapor 1975).

In Carter's (1986) opinion two types of beach ridges may develop on a progradational sea coast. The first type is a result of gradual accretion and c[illegible]ing of swash bars during a deposit's trans[illegible] wave action. This type of beach ridge is constructed by seaward dipping laminae of sand or gravel [illegible]sh laminations. The second type is connected with longshore bar emergence during low wave energy conditions and simultaneous fall of sea level. The morphology of these ridges is more complicated. They are constructed mainly by landward dipping laminae. However, tabular planar cross-lamination connected with landward migration of the lee slope of the emerge feature are also situated here. On the seaward slope of this type of beach ridge a thin layer of swash lamination can be present.

Beach ridges are also partially created by the processes of aeolian deflation and accumulation. There is often an accumulated cover of aeolian deposits on earlier formed ridges, stabilized by vegetation. As a result, on the beach ridges irregular hummock dunes or parallel foredune ridges can be situated. In this case, the sediments of beach ridges are usually separated from aeolian covers by fossil erosional surfaces with a shell or gravel pavement (Carter and Wilson 1990).

The formation of beach ridges is very dependent on conditions of beach supply by littoral deposits. Beach ridges are created only when wave action, and connected with it sea currents, provide more deposits to the beach than the waves can remove (Johnson 1919). Important factors during the creation of beach ridges are the bathymetry of the inner shelf, abundant sediment supply in the nearshore zone and also the wave energy regime and fluctuations of sea level. The average size of a beach's material is also a very important factor. On sandy beaches, the beach ridges accumulate during the low wave energy events, but on gravelly beaches the formation of ridges is usually connected with high wave energy events.

Beach ridges may appear as a single form, as well as a complex of forms, creating often expansive plains of beach ridges. These plains are especially characteristic of progradational coasts. The relief of the individual beach ridges is different. Their height may reach values from a few dozens of centimetres to about a few metres. The distance between two different beach ridges also varies. It is usually thought that the smaller, closely spaced ridges are formed during rapid beach progradation, and that the ridges of larger dimensions and greater spacing are connected with a relatively slower rate of growth (Taylor and Stone 1996).

Beach ridges are good palaeogeography indicators of past wave regimes, sediment supply, sediment source, climatic conditions, sea-level change and also isostatic emergence or submergence of land. If we can measure the absolute age of beach ridges, e.g. using radiocarbon or archaeological methods, we will be able to reconstruct the ancient shorelines' position as well as speed of coast progradation. Beach ridges can also be used to understand past relative sea-level changes and the history of deposit availability and abundance within the inner shelf.

References

Carter, R.W.G. (1986) The morphodynamics of beach ridge formation; Magilligan, Northern Ireland, *Marine Geology* 73, 191–214.

Carter, R.W.G. and Wilson, P. (1990) The geomorphological, ecological and pedological development of coastal foredunes at Magilligan Point, Northern Ireland, in K.F. Nordstrom, N.P. Psuty and R.W.G. Carter (eds) *Coastal Dunes: Form and Process*, 129–157, Chichester: Wiley.

Johnson, D.W. (1919) *Shore Processes and Shoreline Development*, New York: Wiley.

Stapor, F.W. (1975) Holocene beach ridge plain development, Northwest Florida, *Zeitschrift für Geomorphologie, Supplementband* 22, 116–141.

Taylor, M. and Stone, G.W. (1996) Beach-ridges: a review, *Journal of Coastal Research* 12(3), 612–621.

SEE ALSO: beach; beach–dune interaction; beach sediment transport; chenier ridge

RYSZARD K. BORÓWKA

BEACH ROCK

'A consolidated deposit that results from lithification by calcium carbonate of sediment in the intertidal and spray zones of mainly tropical coasts' (Scoffin and Stoddart 1983: 401). Some authors have also used the term 'cay sandstone' to distinguish between rocks that are formed in the supratidal zone and beach rock or beach sandstone formed in the intertidal zone of coral reefs (see, for example, Gischler and Lomando 1997). Although the latitudinal limits of most contemporary beach rocks are approximately 35°N to 35°S, from time to time they do occur in higher latitudes, for example in north-west Scotland (Kneale and Viles 2000).

Beach rock is also widespread on beaches around the Mediterranean Sea (Plate 17) but is perhaps best known for its association with the calcium carbonate beaches of coral reef islands. Beach rock is geomorphologically important in that it preserves coastal landforms, provides a record of former sea levels, creates distinctive pavements and forms very rapidly. It also displays suites of characteristic erosional landforms that include micro-scarps, ridges and runnels and various weathering forms produced by biological processes, chemical erosion and salt attack.

The origin of beach rock has perplexed investigators ever since it was described in the early nineteenth century by travellers like Admiral Beaufort and Charles Darwin. Proposed mechanisms of formation include both physico-chemical and biological models. The former involve cement precipitation resulting from evaporation, CO_2 degassing owing to wave agitation and increasing temperature, and mixing of alkaline fresh water with sea water. Such models tend to have dominated the literature. However, the role of micro-organisms is now being seriously considered as a result of both

Plate 17 Beach rock developed on the south-east coast of Turkey at Arsuz near Iskenderun

field (Webb *et al.* 1999) and laboratory evidence (Neumeier 1999). High Mg calcite (often micritic) and aragonite are the commonest cement types, although low Mg calcite cements are common from temperate zone beach rocks. The cements occur most commonly as clean isopachous fringes of acicular crystals around grains, but meniscus and gravitational cements are also known (Scoffin and Stoddart 1983).

References

Gischler, E. and Lomando, A.J. (1997) Holocene cemented beach deposits in Belize, *Sedimentary Geology* 110, 277–297.

Kneale, D. and Viles, H.A. (2000) Beach cement: incipient $CaCO_3$ – cemented beach rock development in the upper intertidal zone, North Uist, Scotland, *Sedimentary Geology* 132, 165–170.

Neumeier, U. (1999) Experimental modelling of beach rock cementation under microbial influence, *Sedimentary Geology* 126, 35–46.

Scoffin, T.P. and Stoddart, D.R. (1983) Beach rock and intertidal cements, in A.S. Goudie and K. Pye (eds) *Chemical Sediments and Geomorphology*, 401–425, London: Academic Press.

Webb, G.E., Jell, J.S. and Baker, J.C. (1999) Cryptic intertidal microbialites in beach rock, Heron Island, Great Barrier Reef: implications for the origin of microcrystalline beach rock cement, *Sedimentary Geology* 126, 317–334.

A.S. GOUDIE

BEACH SEDIMENT TRANSPORT

Beach sediment transport occurs over the whole area where wave-induced currents are capable of moving sediment: in the shoaling zone, surf zone, breakers and swash zone, and also as aeolian sediment transport on beaches. Spatial gradients in the sediment transport rate determine positions of erosion and deposition and thus three-dimensional changes in the beach shape. Research on beach sediment transport concentrates on predicting the mechanics of transport, direction of transport, transport rate, transport volume, and changes in beach morphology.

The essential difference between sediment transport under a steady unidirectional flow and under waves is that oscillatory BOUNDARY LAYERS show a temporal variation which is not present under steady flow, growing and decaying twice every wave cycle as flows accelerate, decelerate and change direction. Boundary layers are not able to develop fully under oscillatory flow, and consequently are always thinner than under an equivalent unidirectional flow. This means that for a given free-stream velocity and bed roughness, the bed shear stress in oscillatory flow is always larger than beneath steady flow.

Beach sediment transport can be in the form of SUSPENDED LOAD, BEDLOAD and sheet flow (see SHEET EROSION, SHEET FLOW, SHEET WASH). Bedload transport is often modelled as a function of the shear stress acting on the sediment grains, while suspended transport is generally calculated as the product of velocity and concentration profiles. Suspended sediment transport generally receives more attention than bedload transport; however, this is mainly because instruments such as the optical backscatter sensor and acoustic backscatter sensor have been developed which are capable of making high-frequency measurements of suspended sediment concentrations simultaneously with velocity measurements. Because only suspended sediment transport can be measured in this way, many researchers conclude that the most reasonable approach at present is to assume that suspended sediment transport dominates when strong wave motion is present. However, this assumption will remain essentially untested until instruments are developed which are capable of high-frequency measurement of bedload transport.

A number of mechanisms contribute to beach sediment transport, including turbulence, mean currents, currents generated by oscillatory waves at incident and infragravity frequencies, and wave–current interaction. Wave-induced currents include both unidirectional currents (such as longshore currents, RIP CURRENTs and undertow) and rapidly reversing asymmetrical cross-shore currents which flow onshore under the

crest of the wave and offshore under the trough. In combined flows, oscillatory wave motion is generally assumed to entrain sediment which is then moved by a steady current. Beach sediment transport is also affected by local bed slope and the presence of RIPPLES and other bedforms, and can be modified by human activity such as BEACH NOURISHMENT and coastal engineering structures.

The easiest approach to beach sediment transport is to consider cross-shore and longshore transport separately. Longshore transport is responsible for changes in the beach plan shape, and is usually considered to be unidirectional in the direction of the longshore current. In the simplest formulation, the longshore sediment transport rate is proportional to the longshore wave energy flux at the breakpoint. (See LONGSHORE (LITTORAL) DRIFT.)

Cross-shore sediment transport is responsible for changes in the beach profile. This includes features on the subaerial profile such as beach face slope and berms, and submerged features such as nearshore bars. Net cross-shore sediment transport is difficult to calculate because it occurs as an accumulation of small differences between the large values of onshore and offshore transport, which must be evaluated separately and correctly. Most field data and model predictions indicate that offshore sediment transport dominates under breaking waves due to the seaward-directed undertow. During non-breaking wave conditions, transport is generally onshore-directed due to the effects of incident wave asymmetry.

Governing equations based on fundamental physics have not yet been established, and no unified theory of sediment transport presently exists that is valid for all water depths and fluid motions in the nearshore. Instead, there are many sediment transport models, ranging from quasi-steady formulas such as the energetics approach described below to complex numerical models involving higher-order turbulence closure schemes that attempt to resolve the flow field at small scales. Models can be classified by direction (cross-shore or longshore), driving force (e.g. bottom fluid velocity, bed shear stress, wave energy dissipation, stream power, etc.), or mode of transport (bedload, suspended load, total load). Reviews of sediment transport models are given by Schoones and Theron (1995), Bayram *et al.* (2001) and Davies *et al.* (2002), and measurement of coastal sediment transport is reviewed by White (1998).

One of the preferred approaches to modelling both longshore and cross-shore wave and current-induced sediment transport is based on the energetics approach of Bagnold (1963), formulated for a time-varying flow field (e.g. Bailard 1981). The energetics approach assumes that sediment is mobilized by the oscillatory flow under waves and is related to some power of the instantaneous velocity. Once mobilized, sediment can be transported by a number of different mechanisms: time-averaged flows (undertow or longshore currents), asymmetric orbital velocities and gravity in the downslope direction. The fluid forces which drive the energetics model are based on the calculation of various moments of the fluid velocity, which give the direction and magnitude of both oscillatory and mean flows. Use of the energetics model requires knowledge of the moments of the instantaneous flow field, often decomposed into mean, gravity and infragravity band components and then time-averaged. Net transport is calculated from the integral of the instantaneous rate through a particular time interval.

The energetics model is regarded as one of the best theoretical models available at the moment for time-dependent nearshore sediment transport because of its capacity to represent a wide variety of transport conditions in a computationally efficient manner. However, it does not include a number of factors which are believed to be important in beach sediment transport, such as turbulence, fluid accelerations, threshold of motion, and transport over bedforms. In particular, the energetics model may not be appropriate for use in the swash zone, where beach accretion is most likely to occur. Additional processes are likely to be important in swash sediment transport, such as infiltration/exfiltration, bore-generated turbulence, water depth, inertial forces on coarse grains, and sediment advection and convection.

References

Bagnold, R.A. (1963) Mechanics of marine sedimentation, in M.V. Hill (ed.) *The Sea*, Volume 3, 507–528, New York: Wiley.

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

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.

Davies, A.G., van Rijn, L.C., Damgaard, J.S., van de Graaff, J. and Ribberink, J. (2002) Intercomparison

of research and practical sand transport models, *Coastal Engineering* 46, 1–23.
Schoones, J.S. and Theron, A.K. (1995) Evaluation of 10 cross-shore sediment transport/morphological models, *Coastal Engineering* 25, 1–41.
White, T.E. (1998) Status of measurement techniques for coastal sediment transport, *Coastal Engineering* 35, 17–45.

Further reading

Komar, P.D. (1998) *Beach Processes and Sedimentation*, 2nd edition, Englewood Cliffs, NJ: Prentice Hall.
Pethick, J.S. (1984) *An Introduction to Coastal Geomorphology*, London: Edward Arnold.
Short, A. (1999) *Handbook of Beach and Shoreface Morphodynamics*, Chichester: Wiley.

SEE ALSO: bar, coastal; beach; beach–dune interaction; beach nourishment; bedload; current; longshore (littoral) drift; sediment budget; sediment transport; sheet erosion, sheet flow, sheet wash; suspended load; wave

DIANE HORN

BEDFORM

The transport of sand or gravel as BEDLOAD (and see SEDIMENT LOAD AND YIELD) creates on the bed a variety of features the size, form and relative orientation of which depend through complex interactions on the density, shape and coarseness of the sediment particles, and on the strength, uniformity and steadiness of the current. These features are bedforms. They participate in the sediment transport and are normally very small or small in height compared to flow thickness. Although the main kinds are, generally speaking, agreed upon, there is little uniformity as to their nomenclature. Bedforms are known from rivers, tidal systems (especially estuaries) and the deep sea, and are most widely familiar from deserts (see DUNE, AEOLIAN). They contribute significantly to contemporary landscapes and seascapes and, where preserved in Quaternary sediments, are valuable indicators of environment and sediment transport strength and direction. Bedforms and their internal structures can be used to establish sediment-transport directions not only on a regional scale but also locally, where changes in strength and direction have occurred on small geographical and stratigraphical scales. For the river and irrigation engineer, bedforms are among the most important determinants of channel hydraulic ROUGHNESS and resistance to flow.

Extensive field and laboratory studies show that river bedforms and their distinctive patterns of internal stratification can be placed in a number of fields defined more or less closely by grain size and flow strength (Figure 12). At flow strengths below the entrainment threshold (see INITIATION OF MOTION) there is neither sediment transport nor bedforms. As flow strength rises, the bedforms first to appear are current ripples (medium- and finer-grained quartz-density sands) and lower-stage plane beds (coarser sands, granules, gravel). At equilibrium, current ripples are ridges of linguoid plan about 0.02 m high and 0.1–0.2 m in wavelength, which increases with sediment size (see RIPPLE). They move very slowly beneath the current as grains are eroded from the long upstream slope and then deposited by settling and avalanching on the steep leeward face (see REPOSE, ANGLE OF) overlain by a turbulent, recirculating vortex. An internal pattern of cross-lamination records the successive positions of the migrating downstream face. Lower-stage plane beds are underlain by subhorizontal parallel laminae on a millimetre to coarser scale. In sediments coarser than medium-grained sand, but at increasingly large flow strengths for progressively finer grades, lower-stage plane beds and current ripples are replaced by dunes. Like current ripples, these forms are transverse ridges which migrate by the erosion of particles from the upstream side followed by their deposition on the downstream face after settling through and avalanching beneath a recirculating, leeward vortex (see REPOSE, ANGLE OF). Patterns of cross-bedding occur internally, the foresets dipping in the sediment-transport direction. Unlike current ripples, dunes scale on flow depth, varying in height from about one-twentieth to about one-fifth of the depth. In large rivers, such as the Mississippi and Brahmaputra, they are several metres high and one to two hundred metres in wavelength. Extensive fields of even larger dunes, composed of cobble and boulder gravel, have been created by some major catastrophic floods (see DAM; ICE DAM, GLACIER DAM; OUTBURST FLOOD). At large enough flow strengths, current ripples and dunes become increasingly round-crested and flat, and are replaced by upper-stage plane beds over which sediment transport, in the form of very low bed waves, is intense. Internally, forms of subhorizontal laminae and bedding occur within such beds. In the case of sands, the surfaces of the laminae carry faint flow-parallel ridges, called parting or primary current

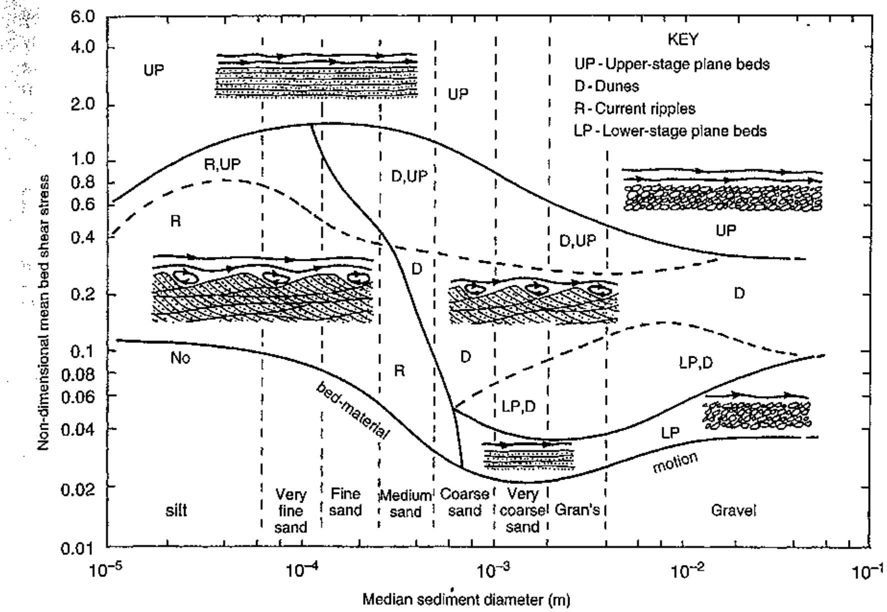

Figure 12 Existence fields, defined by flow strength and median grain diameter, for bedforms and their internal sedimentary structure as shaped by unidirectional water currents in quartz-density sediments. The diagram is based on many hundreds of individual observations. Flow strength is given in a non-dimensional form, defined as the quotient of the mean bed shear stress and the product of the relative particle density, acceleration due to gravity and median particle diameter

lineations, which may be related to flow patterns within the grain-dense lower part of the turbulent BOUNDARY LAYER.

These bedforms are restricted to subcritical flows (see ANTIDUNE), marked by comparatively smooth water surfaces and low values of flow velocity compared to flow depth. Supercritical flows over loose beds consist of unstable, transverse, surface waves more or less in phase with similar waves on the bed, but of a lower height. These are antidune bedforms. At high levels of supercriticality, various very flat bed waves arise, which include rhomboid ripples and dunes related to surface shocks. As supercriticality is promoted by shallow depths, supercritical bedforms can arise at almost any flow strength capable of sediment transport.

Tidal currents, reversing and rotating on a semidiurnal or diurnal scale, are as non-uniform as river flows, but hugely more unsteady, and this factor complicates the shape, internal structure and relationship to flow conditions of the bedforms encountered in estuaries (see ESTUARY) and shallow seas. Additional kinds of bedform are recorded from these environments, as well as those familiar from rivers. Current ripples, sand and gravel dunes, upper-stage plane beds and antidunes are chiefly restricted to the shallower channels and intertidal shoals of estuaries. As an expression of the unsteady conditions, drapes of mud deposited when the water is slack may accumulate in the troughs of the ripples and dunes, and later become preserved within the cross-stratified interiors of the forms. Large bedforms – sand ribbons and sand waves – are found below tide level in the deeper channels and on tide-swept floors of confined seas such as the English Channel, the Southern Bight of the North Sea and Cook Inlet, Alaska. Subtidal sand waves were discovered in the 1920s and 1930s as the result of detailed hydrographic surveys and the appearance of practical echo-sounders. A few decades elapsed before the development of side-scan sonar allowed sand ribbons to be recognized.

Sand ribbons are long, flow-parallel belts of ripple- or dune-covered sand or fine gravel of low relief with a spacing across the current of a few to several times the flow depth. They express bedload transport under conditions of restricted sediment supply. Sand waves are series of long-crested ridges of sediment arranged transversely to the stronger of the tidal currents. The largest, found in the deepest waters, measure 5 m or more in height and a few hundred metres in spacing. They assume a roughly symmetrical, trochoidal profile where flood and ebb tidal streams are comparably strong, the small- to medium-sized dunes on their backs reversing in direction of travel with each change in tidal phase. Sand waves become increasingly asymmetrical in profile as the ebb and flood tidal streams become more unequal in their ability to transport sediment, and the dunes they carry may migrate only in the direction of the stronger flow, although being slightly rounded by the weaker stream. The internal structure of sand waves is not well known but is certainly complex, reflecting the presence of superimposed dunes which may change and reverse their direction of movement as the tidal streams reverse and rotate. Internally, the more symmetrical waves seem to consist of comparatively thin cross-bedded units recording sediment transport in many different but largely opposed directions. The more asymmetrical ones reveal internally a 'master-bedding' that dips gently in the direction of the stronger tidal stream and between which appears to lie cross-bedding facing chiefly in that same direction.

Currents strong enough to transport sand-sized particles in places affect large areas of the ocean floor and the deeper parts of open continental shelves. These variable but essentially unidirectional flows are not of tidal origin but depend on various thermohaline effects. Sand ribbons and transverse structures which have been called sand waves (very large dunes) have been described from many of the places where these currents operate, such as the long and intricate narrows between the Baltic Sea and the North Sea, the ocean floor immediately west of Gibraltar, swept by the Mediterranean Undercurrent, the continental shelf of south-east Africa, affected by the Agulhas Current, and the level tops of several oceanic guyots. The sediments involved are of diverse origins. They range from terrigenous sands, in some cases reworked after being introduced from shallower depths by TURBIDITY CURRENTS, to bioclastic debris (chiefly shells or foraminifera) eroded from adjacent parts of the ocean floor. Other than their location, general form and link with strong currents, little is known or understood about these deep-sea bedforms.

Further reading

Alexander, J., Bridge, J.S., Cheel, R.J. and LeClair, S.F. (2001) Bedforms and associated sedimentary structures formed under supercritical water flows over aggrading sand beds, *Sedimentology* 48, 133–152.

Allen, J.R.L. (1982) *Sedimentary Structures*, Amsterdam: Elsevier.

Allen, J.R.L., Friend, P.F., Lloyd, A. and Wells, H. (1994) Morphodynamics of intertidal dunes: a year-long study at Lifeboat Station Bank, Wells-next-the-Sea, Eastern England, *Philosophical Transactions of the Royal Society of London* A347, 291–345.

Ashley, G.M. (1990) Classification of large-scale subaqueous bedforms: a new look at an old problem, *Journal of Sedimentary Petrology* 60, 160–172.

Berné, S., Castaing, P., Le Drezen, E. and Lericolais, G. (1993) Morphology, internal structure and reversal of asymmetry of large subtidal dunes in the entrance to the Gironde Estuary (France), *Journal of Sedimentary Petrology* 63, 780–793.

Bridge, J.S. and Best, J.L. (1988) Flow, sediment transport and bedform dynamics over the transition from dunes to upper-stage plane beds: implications for the formation of planar laminae, *Sedimentology* 35, 753–763.

Bridge, J.S. and Best, J. (1997) Preservation of planar lamination due to migration of low-relief bed waves over aggrading upper-stage plane beds: comparison of experimental data with theory, *Sedimentology* 44, 253–262.

Carling, P.A. (1999) Subaqueous gravel dunes, *Journal of Sedimentary Research* 69, 534–545.

Carling, P.A., Gölz, E., Orr, H.G. and Redecki-Pawlik, A. (2000) The morphodynamics of fluvial sand dunes in the River Rhine, near Mainz, Germany. I. Sedimentology and morphology, *Sedimentology* 47, 227–252.

Dalrymple, R.W. and Noble, R.N. (1995) Estuarine dunes and bars, in G.M.E. Perillo (ed.) *Geomorphology and Sedimentology of Estuaries*, 359–422, Amsterdam: Elsevier.

Flemming, B.W. (1980) Sand transport and bedform patterns on the continental shelf between Durban and Port Elizabeth (southern African continental margin), *Sedimentary Geology* 26, 179–205.

Kenyon, N.H. and Belderson, R.H. (1973) Bedforms in the Mediterranean undercurrent observed with side-scan sonar, *Sedimentary Geology* 9, 77–99.

Simons, D.B. and Richardson, E.V. (1966) Resistance to flow in alluvial channels, *United States Geological Survey Professional Paper* 422-J, 1–61.

Stride, A.H. (ed.) (1982) *Offshore Tidal Sands*, London: Chapman and Hall.

Weedman, S.D. and Slingerland, R. (1985) Experimental study of sand streaks formed in turbulent boundary layers, *Sedimentology* 32, 133–145.

J.R.L. ALLEN

BEDLOAD

By mode of transport, the sediment load is divided into SUSPENDED LOAD and bedload. The bedload typically consists of coarse particles derived from the bed material. The immersed weight of these particles is supported by a combination of fluid forces and solid reactive forces exerted at intermittent or continuous contacts with the bed (Abbott and Francis 1977). Bedload is dispersed in a zone immediately above the bed surface and is transported in the rolling/sliding or saltating modes. Particles comprising the bedload continually move in and out of storage on the bed. The pattern of particle 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, and each particle spends a negligible time in motion compared to the time spent at rest. Consequently, the virtual velocity of bedload is much lower than the flow velocity; in water, for example, it 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).

In wind, most sand particles move by SALTATION. The saltating particles interact with the bed surface and disturb stationary grains (Anderson and Haff 1991). This not only reduces the threshold of particle motion, it also promotes reptation (the movement of particles impacted by saltating grains over short distances) and surface creep. Within the saltation layer most grains move within 1 to 2 cm of the sand surface. The size of the mobile grains decreases and there is an exponential decline in sediment and mass flux concentration with height (Anderson and Hallet 1986). Accurate data on sediment fluxes are difficult to obtain (Greeley *et al.* 1996; Iversen and Rasmussen 1999), and sand transport equations are frequently used to predict aeolian sand transport rates (Sarre 1989). Above the threshold for sand movement, sand transport rates are commonly assumed to be proportional to the wind (shear) velocity. Aeolian sand transport in bulk is associated with the formation of BEDFORMS of varying size, that develop as regularly repeating patterns (Anderson 1987; Lancaster 1988), and field observation suggests there is close agreement between measured and simulated patterns of erosion and deposition on dunes and wind velocity and direction (Howard *et al.* 1978).

In water, the maximum size of sediment that can be moved by a given flow condition defines flow competence, but the size and amount of sediment moved as bedload is constrained by a river's transport capacity. Transport rates may also be limited by sediment availability. Continuity of bedload transport typically is not maintained along a river, because transport capacity usually does not match the sediment supply. This may promote scour or fill of the river bed and other adjustments to channel geometry. For this reason, although bedload typically constitutes only a few per cent of the total sediment load of most rivers, bedload transport is a very significant process as it governs virtually all aspects of morphological change in river channels. Downstream through a drainage basin, the bed material generally becomes finer through the action of sorting and abrasion; in consequence, the suspended load increasingly dominates over the bedload.

At the lower limit of active transport, where rolling is the dominant transport mode, bed pocket geometry determines which particle sizes are mobile (Andrews 1994). When conditions are below the threshold for general bed motion or the sediment supply is limited the bedload transport rate is moderated by the interaction of coarse and fine size fractions in the bed material, as well as by the available shear stress (Gomez 1995). ARMOURING compensates for the intrinsically lower mobility of coarse particles relative to that of fine grains and renders all particle sizes on the bed surface equally mobile (Parker and Klingeman 1982). Equal mobility arises as a consequence of the shielding of small grains from the flow and the preferential exposure of large particles, coupled with their relative abundance on the bed surface. The adjustments combine to counteract the absolute size effect of particle weight by making coarse particles more available to the flow, and enhancing their probability for entrainment. There are two facets to equal mobility. Equal entrainment mobility is defined as the case when all particle sizes comprising the bed material begin to move at the same flow strength. Equal transport mobility refers to the situation where all particle sizes are transported according to their relative proportions in the bed material, so that the bedload and bed material grain-size distributions are identical. Departures from these conditions give rise to differences in the transport rate of individual size fractions (Wilcock and McArdell 1993), and to hydraulic sorting. Hydraulic sorting is known to occur during the entrainment, transport and deposition of heterogeneous bedload; it is important because of its links to channel armouring and downstream fining (Paola *et al.* 1992).

In most rivers, bedload transport is highly variable in time and space. Temporal variability in bedload transport rates, which is independent of variations in flow conditions, arises from three main sources (Gomez *et al.* 1989). First, variations may result from long- to intermediate-term changes in the rate at which sediment is supplied to or distributed within a channel or reach. Second, short-term, often quasi-cyclic, variations in bedload-transport rates may occur in response to the temporary exhaustion of the supply of transportable material, to the migration of BEDFORMS or groups of particles, or to processes such as ARMOURING. Third, instantaneous fluctuations in bedload transport rates result from the inherently stochastic nature of the physical processes that govern the entrainment and transport of bedload. Spatial variability in bedload transport rates results from downstream and cross-channel changes in the transport field, that occur primarily in response to differences in the shear stress and to changes in the local relation between boundary shear stress and sediment transport (Dietrich and Whiting 1989).

Commonly utilized approaches for gaining knowledge of the bedload transfer through a river reach involve field sampling or measurement, and the application of a formula. Sampling involves the collection of discrete quantities of bedload at various points across a channel, over limited time intervals. Measurement involves the continuous or time-integrated monitoring of bedload over the entire cross-section or reach. The presence of a sampling device on the river bed necessarily alters the pattern of the flow and sediment transport in its vicinity. Thus, bedload samplers must be calibrated to determine their efficiency under different hydraulic and sediment transport conditions. Determining the hydraulic efficiency of a bedload sampler has proved to be a relatively simple task, but determining the sampling efficiency is considerably more complex. Consequently, the sampler calibration process remains incomplete because none of the tests performed to date on any sampling device has provided definitive results (Thomas and Lewis 1993). Since bedload transport rates vary across channel and with time, appropriate temporal and spatial sampling strategies also are required to minimize sampling errors, which decrease as the number of samples collected increases and the number of traverses of the channel over which the samples are collected increases (Gomez and Troutman 1997).

Measurements are usually regarded as exact, and most commonly are obtained using a pit trap. Traps also have a distinct advantage over samplers in as much as, if the trap spans the entire width of the river, it is not only possible to catch all the bedload that passes through the measuring section in a given period of time but also to continuously measure the rate at which sediment accumulates. The simplest traps consist of lined pits or slots in the streambed in which the bedload collects over one or more events (Church *et al.* 1991). More sophisticated traps continuously weigh the mass of sediment (Reid *et al.* 1980).

Bedload formulae equate the rate at which bedload is transported with a specific set of hydraulic and sedimentological variables, and predict bedload transport capacity under given flow conditions. The underlying physics appear quite straightforward (Bagnold 1966), but the conditions governing fluvial bedload transport are complex and there is little consensus about the fundamental hydraulic and sedimentological quantities involved. Consequently, there are numerous bedload transport formulae (Gomez and Church 1989), and none has been universally accepted or recognized as being especially appropriate for practical application.

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.

Anderson, R.S. (1987) A theoretical model for aeolian impact ripples, *Earth Science Reviews* 10, 263–342.

Anderson, R.S. and Haff, P.K. (1991) Wind modification and bed response during saltation of sand in air, *Acta Mechanica Supplement* 1, 21–51.

Anderson, R.S. and Hallet, B. (1986) Sediment transport by winds: toward a general model, *Geological Society of America Bulletin* 97, 523–535.

Andrews, E.D. (1994) Marginal bed load transport in a gravel bed stream, Sagehen Creek, California, *Water Resources Research* 30, 2,241–2,250.

Bagnold, R.A. (1966) An approach to the sediment transport problem from general physics, *US Geological Survey Professional Paper* 422-I.

Church, M.A., Wolcott, J.F. and Fletcher, W.K. (1991) A test of equal mobility in fluvial sediment transport: behaviour of the sand fraction, *Water Resources Research* 27, 2,941–2,951.

Dietrich, W.E. and Whiting, P.J. (1989) Boundary shear stress and sediment transport in river meanders of sand and gravel, in S. Ikeda and G. Parker (eds) *River Meandering*, 1–50, Washington, DC: American Geophysical Union.

Gomez, B. (1995) Bedload transport and changing grain size distributions, in A.M. Gurnell and G. Petts (eds) *Changing River Channels*, 177–199, Chichester: Wiley.

Gomez, B. and Church, M. (1989) An assessment of bedload sediment transport formulae for gravel-bed rivers, *Water Resources Research* 25, 1,161–1,186.

Gomez, B. and Troutman, B.M. (1997) Evaluation of process errors in bed load sampling using a dune model, *Water Resources Research* 33, 2,387–2,398.

Gomez, B., Naff, R.L. and Hubbell, D.W. (1989) Temporal variations in bedload transport rates associated with the migration of bedforms, *Earth Surface Processes and Landforms* 14, 135–156.

Greeley, R., Blumberg, D.G. and Williams, S.H. (1996) Field measurement of the flux and speed of wind-blown sand, *Sedimentology* 43, 41–52.

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.

Howard, A.D., Morton, J.B., Gad-el-Hak, M. and Pierce, D.B. (1978) Sand transport model of barchan dune equilibrium, *Sedimentology* 25, 307–338.

Iversen, J.D. and Rasmussen, K.R. (1999) The effect of wind speed and bed slope on sand transport, *Sedimentology* 46, 723–731.

Lancaster, N. (1988) Controls on aeolian dune size and spacing, *Geology* 16, 972–975.

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

Paola, C., Parker, G., Seal, R., Sinha, S.K., Southard, J.B. and Wilcock, P.R. (1992) Downstream fining by selective deposition in a laboratory flume, *Science* 258, 1,757–1,760.

Reid, I., Layman, J.T. and Frostick, L.E. (1980) The continuous measurement of bedload discharge, *Journal of Hydraulics Research* 18, 243–249.

Sarre, R.D. (1989) Aeolian sand transport, *Progress in Physical Geography* 11, 157–182.

Thomas, R.B. and Lewis, J. (1993) A new model for bed-load sampler calibration to replace the probability-matching method, *Water Resources Research* 29, 583–597.

Wilcock, P.R. and McArdell, B.W. (1993) Surface-based fractional transport rates: mobilization thresholds and partial transport of a sand-gravel sediment, *Water Resources Research* 29, 1,297–1,312.

BASIL GOMEZ

BEDROCK CHANNEL

Bedrock channels are those with frequent exposures of bedrock in their bed and banks. More precisely, these channels lack a coherent cover of alluvial sediments, even at low flow, although a thin and patchy alluvial cover may be present. However, short-term pulses of rapid sediment delivery may produce temporary sediment fills (see SEDIMENT ROUTING; SEDIMENT WAVE). Bedrock channels exist only where transport capacity exceeds sediment supply over the long term. Contrary to classical definitions, bedrock channels are self-formed. Bed and banks are not composed of transportable sediment, but are erodible. Flow, sediment flux and base-level conditions (see BASE LEVEL) dictate self-adjusted combinations of channel gradient, width and bed morphology.

Bedrock channels are important because (1) they set much of the RELIEF structure of unglaciated mountain ranges, and (2) the controls on their incision rates largely dictate the relationships among climate, lithology, tectonics and topography. Moreover, because river incision rate sets the boundary condition for hillslope EROSION, regional DENUDATION rates and patterns are dictated by the bedrock river network (see HILLSLOPE-CHANNEL COUPLING). Finally, bedrock rivers transmit signals of base level (tectonic/eustatic (see EUSTASY)) and climate change through landscape, and therefore set the timescale of response to perturbation.

Similar to alluvial channels, the longitudinal profiles of bedrock channels are typically smoothly concave-up (see LONG PROFILE, RIVER). These profiles are well described by Flint's law relating local channel gradient (S) to upstream drainage area (A):

$$S = k_s A^{-\theta} \quad (1)$$

where k_s is the steepness index and θ the concavity index. Steepness index is known to be a function of rock uplift rate, lithology and climate (see GRADE, CONCEPT OF). The concavity index is typically in the range 0.4–0.6, and is apparently insensitive to differences in uplift rate, lithology and climate where these are uniform within a DRAINAGE BASIN. However, θ does vary beyond this typical range, usually where downstream fining is particularly strong, or where lithology or uplift rate vary systematically downstream.

Bedrock channel width also varies with drainage area in a manner similar to that observed in alluvial channels (see HYDRAULIC GEOMETRY):

$$W \propto A^{0.3-0.4} \quad (2)$$

Bed morphology also appears to be dynamically adjusted to hydraulic and sediment-flux conditions, and in bedrock-dominated reaches includes STEP-POOL SYSTEMS, plane bed and incised inner channel forms. Discrete KNICKPOINTS and erosional forms such as flutes, POT-HOLES, longitudinal grooves and undulating canyon walls are common.

Processes of erosion in bedrock channels include plucking, macro-abrasion, wear, chemical and physical WEATHERING, and possibly CAVITATION (see CORROSION; FROST AND FROST WEATHERING). These processes all include critical thresholds (see THRESHOLD, GEOMORPHIC) and most work is probably done by large storms (see INITIATION OF MOTION; MAGNITUDE–FREQUENCY CONCEPT). The relative roles of extraction of joint blocks (plucking plus macro-abrasion) and incremental wear are debated, but appear to depend on properties of the substrate lithology and flow conditions (see ROCK MASS STRENGTH). The relative contributions of BEDLOAD and SUSPENDED LOAD to ABRASION (macro- and wear) are also debated, but most agree sediment flux plays a dual role: providing tools for abrasion, but protecting the bed when overly abundant. The exact nature of the dependence of incision rate on sediment flux and grain size, and the different mechanics of plucking, macro-abrasion and wear all have far-reaching consequences for the relations among climate, tectonics and topography. Both DEBRIS FLOWS and FLOODS likely contribute to bedrock channel erosion in mountainous areas. Their relative importance is not well known, but apparently depends on position in the landscape and setting (tectonics, lithology and climate).

Rates of incision of bedrock rivers are highly variable (from mMa^{-1} to cmyr^{-1}), and depend primarily on tectonic setting and other controls on base-level fall. Where they have both been measured, long-term bedrock river incision rates match the highest rock uplift rates. Burbank *et al.* (1996) estimated incision rates up to $12\,\mathrm{mmyr}^{-1}$ on the basis of cosmogenic exposure ages of strath terraces along the Indus River in north-west Pakistan (see COSMOGENIC DATING; TERRACE, RIVER). Short-term incision rates up to $10\,\mathrm{cmyr}^{-1}$ have been measured under extreme circumstances.

Incision rates are positively correlated with channel gradient and drainage area, and are often modelled as a function of bed shear stress. The best known, semi-successful, bedrock river incision model is the shear stress or unit STREAM POWER model:

$$E = KA^m S^n \quad (3)$$

where E denotes vertical incision rate (L/T), A upstream drainage area (L^2), S channel gradient, K is a dimensional coefficient of erosion (L^{1-2m}/T) (see EROSIVITY), and m and n are positive constants that depend on erosion process, channel hydraulics and basin hydrology. Although this simple model has been useful for exploration of interactions among erosion, topography, climate and tectonics, much uncertainty remains regarding the physical controls on the model parameters K, m and n. In addition, equation (3) neglects an incision threshold and therefore misses an important aspect of the physics. Further field and laboratory studies are needed to resolve important outstanding issues.

Reference

Burbank, D.W., Leland, J., Fielding, E., Anderson, R.S., Brozovic, N., Reid, M.R. and Duncan, C. (1996) Bedrock incision, rock uplift and threshold hillslopes in the northwestern Himalayas, *Nature* 379, 505–510.

Further reading

Howard, A.D., Seidl, M.A. and Dietrich, W.E. (1994) Modeling fluvial erosion on regional to continental scales, *Journal of Geophysical Research* 99, 13,971–13,986.

Sklar, L.S. and Dietrich, W.E. (2001) Sediment and rock strength controls on river incision into bedrock, *Geology* 29, 1,087–1,090.

Stock, J.D. and Montgomery, D.R. (1999) Geologic constraints on bedrock river incision using the stream power law, *Journal of Geophysical Research* 104, 4,983–4,993.

Tinkler, K. and Wohl, E.E. (eds) (1998) *Rivers over Rock: Fluvial Processes in Bedrock Channels*, Washington, DC: AGU Press.

Whipple, K.X., Hancock, G.S. and Anderson, R.S. (2000) River incision into bedrock: mechanics and relative efficacy of plucking, abrasion, and cavitation, *Geological Society of America Bulletin* 112, 490–503.

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* 104, 17,661–17,674.

SEE ALSO: channel, alluvial; dynamic equilibrium; palaeoflood; tectonic geomorphology; valley

KELIN X. WHIPPLE

BEHEADED VALLEY

Fluvial valleys running across an active strike-slip fault react to lateral movement along the fault in the way that their lower reaches, located downstream from the fault, become horizontally displaced in relation to the upper reaches, situated upstream from the fault. In this way the continuity of the valley is lost and the lower reach becomes beheaded. Streams may deflect at the fault and follow the fault zone until they turn into the displaced lower reach, or abandon the original valley and continue without deflection. In the latter case the beheaded section of a valley becomes dry. It is usually only small, narrow valleys occupied by minor streams which become beheaded. For larger rivers, floodplains are normally sufficiently wide to retain spatial continuity.

If a beheaded valley can be clearly defined in the field and contains an alluvial suite which can be dated, then slip rate along the fault can be determined.

Further reading

Keller, E.A., Bonkowski, M.S., Korsch, R.J., Shlemon, R.J. (1982) Tectonic geomorphology of the San Andreas fault zone in the southern Indio Hills, Coachella Valley, California, *Geological Society of America Bulletin* 93, 46–56.

SEE ALSO: seismotectonic geomorphology; tectonic geomorphology

PIOTR MIGOŃ

BERGSCHRUND

Deep, transverse or extensional crevasses that occur at the heads of valley or cirque glaciers (see CIRQUE, GLACIAL) are called bergschrunds. They differ from randklufts in that they occur in glacier ice rather than between the ice and the bedrock headwall. The randkluft of a glacier exists due to a combination of preferential ablation adjacent to warm rock surfaces and ice movement away from the rock wall. Both types of crevasse form formidable barriers for climbers in glacierized mountainous terrain and are particularly dangerous when covered in snow. Although numerous studies have suggested that the bergschrund of a glacier separates immobile, cold based ice at the head of a glacier from the active, sliding ice lower down, Mair and Kuhn (1994) have demonstrated that ice was sliding both above and below the position of a bergschrund in a glacier in the Austrian Alps. Early work on bergschrunds suggested that they were the location of intense FREEZE–THAW CYCLE activity and were, therefore, crucial to the excavation of cirques. Several problems with this hypothesis became apparent once bergschrunds were visited more frequently, most notably by W.R.B. Battle. Essentially, the base of bergschrunds, where bedrock is only occasionally encountered, do not experience appreciable freeze–thaw cycles. It is now accepted that the most effective conditions for freeze–thaw activity lie in the randkluft rather than in the bergschrund of a glacier (Gardner 1987).

References

Gardner, J.S. (1987) Evidence for headwall weathering zones, Boundary Glacier, Canadian Rocky Mountains, *Journal of Glaciology* 33, 60–67.

Mair, R. and Kuhn, M. (1994) Temperature and movement measurements at a bergschrund, *Journal of Glaciology* 40, 561–565.

Further reading

Battle, W.R.B. and Lewis, W.V. (1951) Temperature observations in bergschrunds and their relationship to cirque erosion, *Journal of Geology* 59, 537–545.

Embleton, C. and King, C.A.M. (1975) *Glacial Geomorphology*, London: Edward Arnold.

Lewis, W.V. (ed.) (1960) *Norwegian Cirque Glaciers*, Royal Geographical Society Research Series No. 4.

Thompson, H.R. and Bonnlander, B.H. (1956) Temperature measurements at a cirque bergschrund in Baffin Island: some results of W.R.B. Battle's work in 1953, *Journal of Glaciology* 2, 762–769.

DAVID J.A. EVANS

BIOGEOMORPHOLOGY

Biogeomorphology is sometimes used as an umbrella term to describe studies which focus on the linkages between ecology and geomorphology. Because biogeomorphology deals with the interface between two disciplines it is necessarily diverse, interdisciplinary and hard to define in detail. Biogeomorphological studies have a long history, with several nineteenth-century workers focusing on the interrelationships between communities and landscapes at a range of scales, although the term itself was only coined in the late twentieth century (Viles 1990). Amongst nineteenth-century pioneers, Charles Lyell in 1835 noted the importance of the agency of organic beings in causing superficial modifications on the Earth's surface, and Charles Darwin undertook a classic piece of work on the role of earthworms in influencing soils (Darwin 1881). In recent years several volumes of collected papers on biogeomorphology have been published which provide varying pictures of the scope and nature of biogeomorphological research (see, for example, Viles 1988; Thornes 1990; Hupp *et al.* 1995; Viles and Naylor 2002). Papers within these volumes cover a whole range of organism: geomorphology interactions in riparian, hillslope and coastal settings in environments ranging from arid to humid tropical.

Similar terms have also been used in the literature, including zoogeomorphology or the interrelationship between animals and geomorphology (Butler 1995) and phytogeomorphology (Howard and Mitchell 1985) which investigates the influence of topography on plant communities. Furthermore, geoecology is a commonly used term, especially in the European literature, which also encompasses work addressing interactions between ecology and geomorphology (often at a large scale). Dendrogeomorphology, or the use of tree ring and allied evidence to study geomorphic processes, makes use of the influence of geomorphic processes on plant growth to throw light on the nature and timing of those processes – a neat way of linking ecology and geomorphology from a rather different perspective. Biogeomorphology and these other, similar, umbrella terms reflect a recent trend within the Earth and environmental sciences to investigate links between biotic and abiotic processes (as shown by the flowering of biogeochemistry as a study area, and the growth of interest in Gaia).

Three common themes within present-day biogeomorphological research are the effects of organisms on geomorphic processes, the contribution made by organic processes to the development of landforms, and the impact of geomorphological processes on ecological community development. Many studies have been made in recent years within these themes. For example, in terms of the impact of organisms on geomorphic processes studies have been made of the role of isopods and other fauna in sediment movement in the Negev desert by Yair (1995); the role of *Sabellaria alveolata* reefs in storing coastal sand on the Welsh coast by Naylor and Viles (2000), and the role of plants in influencing splash erosion in Mediterranean matorral environments by Bochet *et al.* (2002). Examples of studies of the role of organic processes in landform development include the study of Fiol *et al.* (1996) which investigates the role of biological weathering in the creation of solutional rillenkarren, and the work of Whitford and Kay (1999) on the role of mammal bioturbation in the production of long-lived mound structures (often called mima mounds). Investigations of the influence of geomorphic processes on ecosystems have been carried out by many ecologists and geomorphologists, such as Scatena and Lugo (1995) in subtropical forests and Hayden *et al.* (1995) on coastal barrier islands. Overall, research into these three major biogeomorphological themes is characterized to date by being largely empirical, field based and focused on a limited range of interactions. There are clear links between the three themes, as for example mammal burrowing produces mima mounds which then influence subsequent vegetation patterning.

Geomorphology and ecology are linked in detail in a range of different ways and understanding and measuring these links has provided much work for biogeomorphologists. Looking at the impact of ecology on geomorphology, organisms can have passive and/or active impacts on geomorphological processes. For example, a micro-organic biofilm can produce chemical weathering of the underlying rock (an active link) whilst also retarding the action of other weathering processes (passively). Biological impacts of geomorphological processes are often referred to by specific terms such as bioerosion, bioweathering, bioturbation, bioconstruction and bioprotection (see Naylor *et al.* 2002 for further details). Considerable research effort has gone into defining these terms and developing ways of studying and quantifying these processes. For example, bioerosion of coastal rocks by a suite of sessile and motile organisms has necessitated measurement of burrow dimensions and calculation of ages of the organisms creating them, as well as quantification of grazing trails through measurement of faecal contents. On the other side of the equation, geomorphology can exert an active and/or passive control on ecosystems. For example, topography influences microclimate which in turn affects plant communities (a passive geomorphological impact) whilst geomorphic processes such as mudflows and rockfalls provide an active control on vegetation development. A whole host of different techniques have been developed to study such influences, often involving mapping and correlation.

All exogenetic geomorphological processes have the potential to be influenced by biological activity; even in some quite hostile environments, as work on subglacial bacterial involvement in chemical weathering has demonstrated (Sharp *et al.* 1999). Indeed, there have been some suggestions that the harsher the environment the more closely interlinked biotic and geomorphic processes are, as organisms extract nutrients, shelter and water from sediments and rocks (Viles 1995). The whole spectrum of biological life forms is involved in biogeomorphological interactions, with animals, plants and a host of micro-organisms all recorded as influencing geomorphic processes. Bacteria have been found to contribute to the precipitation of sinter and travertine in hot spring environments, for example, and tree roots commonly enhance riverbank stability, whilst beaver dams have been recorded as having major impacts on some river networks. As a general rule, micro-organic and plant impacts are more widespread and important to geomorphology than those of animals, which tend to be spatially and temporally patchy in occurrence. Biogeomorphic interactions range greatly in scale and complexity: from the impact of one single organism on rock weathering at the microscale to the involvement of dynamic forest communities in whole catchments. One of the biggest challenges awaiting biogeomorphology in the future is to develop further studies of large-scale ecosystem: geomorphological system interactions over hundreds to thousands of year timescales.

Biogeomorphology is not simply an esoteric scientific backwater dealing with a few bizarre processes (although there are some notably weird examples of biogeomorphic studies such as the work of Splettstoesser in 1985 which discusses the role of rockhopper penguin (*Eudyptes crestatus*) feet in sandstone weathering); it has many applications. Identification of current biological:geomorphological linkages can help geologists interpret unusual sedimentary structures. Recognition of distinctive signatures of biogenic contributions to geomorphic processes on Earth can help scientists search for evidence of former life on other planets such as Mars. More practically, environmental engineering can harness the protective role of organisms in many environments to retard the action of some geomorphological processes. For example, stabilization of coastal dunes through revegetation is an essentially biogeomorphological project. At the smaller scale, understanding links between biofilms and rock surface weathering can aid the conservation of cultural heritage through reducing the threat of biodeterioration.

The future development of biogeomorphology depends both upon its capacity to answer fundamental questions and its ability to provide practical solutions to environmental problems. In some areas, such as the riparian environment, biogeomorphological studies are blossoming and providing much practical information on the mechanical role of roots, the influence of fluvial processes on seed banks and the biochemical role of riparian vegetation. In other areas, biogeomorphological studies remain more narrowly focused on unusual links between single organisms and one geomorphic process. In order to prosper further biogeomorphological studies need to establish novel research methodologies and techniques to investigate the varied links between the biotic and geomorphic worlds, many of which have proved quite hard to quantify and monitor. Furthermore, biogeomorphic studies need to move away from simple empirical, short-term studies to looking at longer term and larger scale situations. For this, numerical modelling may provide a way forward. Also, biogeomorphic studies must try and encompass the evolving two-way interplay between geomorphic and ecological processes, rather than simply focus on the impact of organisms on geomorphic processes or the influence of geomorphology on ecosystem development. Finally, biogeomorphological studies need to continue and expand their essential bridging role – by considering the links between a whole host of organic and inorganic processes in a wide range of environments within a broadly defined Earth surface systems science. The term biogeomorphology is far less important than the scientific terrain it describes – one part of the fertile, dynamic, boundary between the inorganic and organic worlds.

References

Bochet, E., Poesen, J. and Rubio, J.L. (2002) Influence of plant morphology on splash erosion in a Mediterranean matorral, *Zeitschrift für Geomorphologie* 46, 223–243.

Butler, D.R. (1995) *Zoogeomorphology: Animals as Geomorphic Agents*, Cambridge: Cambridge University Press.

Darwin, C. (1881) *Vegetable Mould and Earthworms*, London: John Murray.

Fiol, L., Fornos, J.-J. and Gines, A. (1996) Effects of biokarstic processes on the evolution of solutional

rillenkarren in limestone rocks, *Earth Surface Processes and Landforms* 21, 447–452.
Hayden, B.P., Santos, M.C.F.V., Shao, G. and Kochel, R.C. (1995) Geomorphological controls on coastal vegetation at the Virginia Coast Reserve, *Geomorphology* 13, 283–300.
Howard, J.A. and Mitchell, C.W. (1985) *Phytogeomorphology*, New York: Wiley.
Hupp, C.R., Osterkamp, W.R. and Howard, A.D. (eds) (1995) Biogeomorphology, terrestrial and freshwater systems, *Geomorphology Special Issue* 13, 1–347.
Lyell, C. (1835) *Principles of Geology*, 4th edition, London: John Murray.
Naylor, L.A. and Viles, H.A. (2000) A temperate reef builder: an evaluation of the growth, morphology and composition of *Sabellaria alveolata* (L.) colonies on carbonate platforms in South Wales, in E. Insalaco, P.W. Skelton and T.J. Lamer (eds) *Carbonate Platform Systems: Components and Interactions*, Geological Society Special Publication No. 178, 9–19.
Naylor, L.A., Viles, H.A. and Carter, N.E.A. (2002) Biogeomorphology revisited: looking towards the future, *Geomorphology* 47, 3–14.
Scatena, F.N. and Lugo, A.E. (1995) Geomorphology, disturbance, and the soil and vegetation of two subtropical steepland watersheds of Puerto Rico, *Geomorphology* 13, 199–213.
Sharp, M., Parkes, J., Cragg, B., Fairchild, I.J., Lamb, H. and Tranter, M. (1999) Widespread bacterial populations at glacier beds and their relationship to rock weathering and carbon cycling, *Geology* 27, 107–110.
Splettstoesser, J.F. (1985) Note on rock striations caused by penguin feet, Falkland Islands, *Arctic and Alpine Research* 17, 107–111.
Thornes, J.B. (ed.) (1990) *Vegetation and Erosion: Processes and Environments*, Chichester: Wiley.
Viles, H.A. (ed.) (1988) *Biogeomorphology*, Oxford: Blackwell.
——(1990) 'The agency of organic beings': A selective review of recent work in biogeomorphology, in J.B. Thornes (ed.) *Vegetation and Erosion*, 5–24, Chichester: Wiley.
——(1995) Ecological perspectives on rock surface weathering: towards a conceptual model, *Geomorphology* 13, 21–35.
Viles, H.A. and Naylor, L.A. (2002) Biogeomorphology Special Issue. *Geomorphology* 47, 1, 1–94.
Whitford, W.G. and Kay, F.R. (1999) Biopedturbation by mammals in deserts: a review, *Journal of Arid Environments* 41, 203–230.
Yair, A. (1995) Short- and long-term effects of bioturbation on soil erosion, water resources and soil development in an arid environment, *Geomorphology* 13, 87–100.

HEATHER A. VILES

BIOKARST

Biokarst refers to karst landforms created, or influenced to a significant degree, by biological processes. In turn, the processes involved in the formation of such landforms are often called biokarstic. Biokarst features can be erosional or depositional, or involve a combination of the two processes, and are commonly found on exposed limestone surfaces in a range of environmental settings. An early paper by Jones (1965) described many of the erosional features found on limestone pavements as being at least partly biokarstic in origin. Some TUFAS AND TRAVERTINES are largely influenced by biological processes and thus can be seen to be biokarstic, as can some organically influenced cave deposits. Most landforms recognized as biokarst are quite small (maximum of tens of metres in dimensions), but there is an indirect biokarstic element to most karst landscapes as organisms play a key influence on soil acidity and CO_2 levels which in turn are a vital control of karst development.

Similar terms in the literature include phytokarst, which is more narrowly defined as karst landforms produced by the action of plants, and zookarst, which refers to features produced by animal action. Both phytokarst and zookarst are subsumed within biokarst which can be produced by animal, plant or micro-organism action (and commonly involves a combination of organisms). The classic phytokarst landscape is that described by Folk *et al.* (1973) at Hell, Grand Cayman Island. Here, a series of limestone pinnacles in a low-lying swampy environment have been blackened and dissected in a random spongework pattern which Folk *et al.* ascribe to the action of cyanobacteria (blue-green algae). Another commonly identified type of phytokarst are the light-oriented erosional pinnacles found in the lit zone of many cave entrances (as reported by Bull and Laverty in 1982 in Mulu, Borneo, for example, and sometimes given the alternative name of photokarren). Other phytokarst features are the root holes produced in many limestone surfaces. Zookarstic features are rather rare and localized, but include small-scale erosional relief produced by rock wallaby urine in parts of Australia, and grooves produced by the giant tortoise (*Geochelone gigantea*) on Aldabra Atoll, Indian Ocean. By far the most important group of organisms contributing to biokarstic processes are micro-organisms and lower plants, which in mixed biofilm communities coat most subaerial limestone surfaces in a wide range of environments. Such biofilms play a range of active and passive roles in geochemical transformations, aiding both solution and re-precipitation of calcite.

Biokarst has been recorded from most karst areas, with many studies emanating from the great Chinese karst landscapes. Spectacular biologically influenced erosional relief is also found on many coastal limestone platforms, where bioerosion by a range of organisms produces a complex coastal biokarst. Although karst scientists have often been quick to note biokarst features, it has proved difficult to provide convincing process-form links in order to identify the exact nature and importance of biological influences. The major reason for this difficulty is the multi-factorial nature of karst development, which makes it impossible to untangle the interaction of interlinked processes and emerging forms. Some progress has been made with experimental studies, for example the work of Fiol *et al.* (1996) on the influence of rock surface micro-organism communities in rillenkarren development and the work of Moses and Smith (1993) on the role of physical weathering by lichens on kamenitza evolution.

The small-scale nature of many biokarstic features and the difficulties of positively ascribing their genesis to specific biological processes has made several karst scientists doubt their importance either to karst landscape development or as diagnostic landforms. The most important goal for future work is to provide a more general explanation of the role of a whole range of organisms and biological processes in karst landscape development as a whole, rather than worry whether any individual landform can be defined as biokarstic.

References

Bull, P.A and Laverty, M. (1982) Observations on phytokarst, *Zeitschrift für Geomorphologie* 26, 437–457.
Fiol, L., Fornos, J.J. and Gines, A. (1996) Effects of biokarstic processes on the development of solutional rillenkarren in limestone rocks, *Earth Surface Processes and Landforms* 21, 447–452.
Folk, R.L., Roberts, H.H. and Moore, C.H. (1973) Black phytokarst from Hell, Cayman Islands, British West Indies, *Bulletin, Geological Society of America* 84, 2,351–2,360.
Jones, R.J. (1965) Aspects of the biological weathering of limestone pavements, *Proceedings, Geologists' Association* 76, 421–433.
Moses, C.A. and Smith, B.J. (1993) A note on the role of *Collema auriforma* in solution basin development on a Carboniferous limestone substrate, *Earth Surface Processes and Landforms* 18, 363–368.

Further reading

Viles, H.A. (1984) Biokarst: review and prospect, *Progress in Physical Geography* 8, 523–542.
Viles, H.A. (1988) Organisms and karst geomorphology, in H.A. Viles (ed.) *Biogeomorphology*, 319–350, Oxford: Blackwell.

HEATHER A. VILES

BLIND VALLEY

This is a valley formed by fluvial processes that terminates downstream against a steep and sometimes precipitous slope at the foot of which the stream that carved the valley disappears underground into a cave system. The headwaters are usually on relatively impervious rocks such as sandstones or granites and the surface stream disappears underground when it crosses a lithological contact onto a KARST rock such as limestone. The larger the stream, the further it penetrates into the karst before sinking underground, and hence the longer the associated blind valley. In the early stages of development of blind valleys, the downstream wall is not very steep or high, so if the capacity of the stream-sink (swallow hole, ponore) is exceeded during flood the excess water will overflow downstream along its former course, which is usually dry and abandoned. Such cases are referred to as semi-blind valleys. The incision of the blind valley is controlled by the rate of lowering of the cave system into which it drains. This can proceed in stages as the cave stream breaks through to lower levels. Incision is propagated upstream into the blind valley and results in stream terraces. Thus terraces are often found in blind valleys that grade to the position of a former stream-sink in the terminal face high above the modern swallow hole. Over 10^4 to 10^5 years blind valley incision can attain tens to hundreds of metres.

PAUL W. WILLIAMS

BLOCKFIELD AND BLOCKSTREAM

The term blockfield (or block field) is used to describe an extensive cover of coarse rubble on flat or gently sloping terrain, with an absence of fine material at the ground surface. The German term *felsenmeer* ('stone sea') is sometimes used to describe the same phenomenon. Three types of blockfield are recognized: autochthonous blockfields, formed *in situ* by WEATHERING of the underlying bedrock; para-autochthonous blockfields, in which boulders produced by weathering of bedrock have undergone downslope mass movement over low gradients; and allochthonous blockfields derived from GLACIAL DEPOSITION by upfreezing of boulders and washing out of fine sediments. Blockstreams (or block streams) are covers of coarse debris that have accumulated by mass movement on valley floors.

Most blockfields and blockstreams occur in areas of present or former periglacial conditions (see PERIGLACIAL GEOMORPHOLOGY), particularly in arctic environments and on mid-latitude mountains that lay in the periglacial zone outside the limits of the last Pleistocene ice sheets. Blockfields are particularly widespread on mid- and high-latitude plateaux such as those of Scandinavia and Scotland.

Blockfields and blockstreams occur on a wide range of rock types, but are particularly common on well-jointed igneous and metamorphic rocks that have weathered to produce abundant boulders but only limited amounts of fine sediment. Most blockfields comprise boulders less than 1–2 m in length. In autochthonous blockfields the largest boulders usually occur at the surface and boulder size diminishes with depth. Below the openwork surface layer, blockfields and blockstreams usually contain an infill or matrix of fine sediment (sand, silt and clay), and interstitial organic material has also been recorded. Plateau blockfields tend to be 0.5–4.0 m deep, but blockstreams consisting of accumulated valley-floor boulder deposits reach depths of 10 m or more.

Surface boulders in blockfields may be angular or, more commonly, edge-rounded by GRANULAR DISINTEGRATION. Where downslope mass movement has occurred, elongate boulders often exhibit preferred downslope orientation and upslope imbrication. PATTERNED GROUND may be present in the form of large sorted circles on level ground and sorted stripes on slopes, and blockstreams sometimes support lobate structures indicative of movement by SOLIFLUCTION of underlying fine sediments.

In a perceptive early (1906) account of blockfields and blockstreams on the Falkland Islands, J.G. Andersson attributed their formation to frost weathering (see FROST AND FROST WEATHERING) of the underlying bedrock, slow downslope movement of the weathered debris by solifluction, and immobilization by eluviation (see ELUVIUM AND ELUVIATION) of fine sediment from the upper layers. Upheaving of boulders and frost-sorting also appear necessary to produce downward fining of the openwork boulder layer and the formation of sorted patterned ground. Although this general model is widely accepted, some researchers have suggested that autochthonous and para-autochthonous blockfields and blockstreams are of polygenetic origin. In particular, it has been proposed that some plateau blockfields evolved from chemically-weathered (see CHEMICAL WEATHERING) REGOLITH mantles, of interglacial or Tertiary age, that were subsequently modified by frost action (e.g. Nesje 1989; Rea *et al.* 1996; Dredge 2000). This view is based on the location of blockfields on Tertiary erosion surfaces, and the presence in the subsurface fine fraction of clay minerals indicative of prolonged chemical weathering. On certain lithologies, however, blockfields have developed on glacially eroded bedrock since the last glacial maximum (Ballantyne 1998) implying formation under periglacial conditions alone within the last 20,000 years. Some blockfields also show evidence for modification by glacier ice or glacial meltwater (Dredge 2000).

Although there is evidence for blockfield formation during the Holocene in arctic permafrost environments (Dredge 1992), mid-latitude blockfields and blockstreams are manifestly relict. Exposed boulder surfaces have been edge-rounded by prolonged granular disintegration and many support a cover of mosses and lichens. The relationship between such relict blockfields and former ICE SHEETS has been vigorously debated. In some areas, such as western Norway and north-west Scotland, the lower limits of autochthonous blockfields descend regularly along former glacier flow-lines and have been interpreted as trimlines (see TRIMLINE, GLACIAL) marking the maximum vertical extent of the last ice sheets in these areas (Nesje 1989; Nesje and Dahl 1990; Ballantyne *et al.* 1998). Elsewhere, however, there is convincing evidence that blockfields survived the last glacial maximum under a cover of cold-based glacier ice that was frozen to the underlying substrate and hence accomplished little or no erosion (Kleman and Borgström 1990; Dredge 2000). Thus not only does the age and evolution of blockfields and blockstreams vary from area to area, but also their significance in relation to the dimensions of former ice sheets is dependent on the thermal regime of these ice masses.

References

Andersson, J.G. (1906) Solifluction, a component of subaerial denudation, *Journal of Geology* 14, 91–112.

Ballantyne, C.K. (1998) Age and significance of mountain-top detritus, *Permafrost and Periglacial Processes* 9, 327–345.

Ballantyne, C.K., McCarroll, D., Nesje, A., Dahl, S.O. and Stone, J.O. (1998) The last ice sheet in North-West Scotland: reconstruction and implications, *Quaternary Science Reviews* 17, 1,149–1,184.

Dredge, L.A. (1992) Breakup of limestone bedrock by frost shattering and chemical weathering, eastern Canadian arctic, *Arctic and Alpine Research* 24, 314–323.

——(2000) Age and origin of upland blockfields on the Melville Peninsula, Eastern Canadian Arctic, *Geografiska Annaler* 82A, 443–454.

Kleman, J. and Borgström, I. (1990) The boulderfields of Mt Fulfjället, west-central Sweden, *Geografiska Annaler* 72A, 63–78.

Nesje, A. (1989) The geographical and altitudinal distribution of block fields in southern Norway and its significance to the Pleistocene ice sheets, *Zeitschrift für Geomorphologie, Supplementband* 72, 41–53.

Nesje, A. and Dahl, S.O. (1990) Autochthonous block fields in southern Norway: implications for the geometry, thickness and isostatic loading of the Late Weichselian Scandinavian ice sheet, *Journal of Quaternary Science* 5, 225–234.

Rea, B.R., Whalley, W.B., Rainey, M.M. and Gordon, J.E. (1996) Blockfields, old or new? Evidence and implications from some plateaus in northern Norway, *Geomorphology* 15, 109–121.

Further reading

Ballantyne, C.K. and Harris, C. (1994) *The Periglaciation of Great Britain*, Cambridge: Cambridge University Press.

SEE ALSO: frost and frost weathering; frost heave; mechanical weathering; periglacial geomorphology

COLIN K. BALLANTYNE

BLOWHOLE

Fountains of spray are emitted through blowholes during storms and high tidal periods when large breakers surge into tunnel-like caves connected to the surface. Many blowholes develop along joint (see JOINTING) or fault-controlled shafts, but particularly spectacular examples result from marine invasion of KARST tunnels and sinkholes in limestone regions, and lava tubes or tunnels in volcanic areas. Blowholes are also common on CORAL REEFS, where encrusting coralline algae can enclose spur and groove systems and surge channels running through algal ridges.

Further reading

Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Oxford University Press.

ALAN TRENHAILE

BLUE HOLE

Likened to sapphires set in turquoise, they are submarine, circular, steep-sided holes which occur in coral reefs.

The classic examples come from the Bahamas (Dill 1977), but other instances are known from Belize and the Great Barrier Reef of Australia (Backshall *et al.* 1979). Although volcanicity and meteorite impact have both been proposed as mechanisms of formation, the most favoured view is that they are the product of karstic processes (i.e. they are a DOLINE or CENOTE) which acted at times of low glacial sea levels when the reefs were exposed to subaerial processes. Subsequently they were submerged by the Flandrian Transgression of the Holocene.

References

Backshall, D.G., Barnett, J. and Davies, P.J. (1979) Drowned dolines – the blue holes of the Pompey Reefs, Great Barrier Reef, BMR *Journal of Australian Geology and Geophysics* 4, 99–109.

Dill, R.F. (1977) The blue holes – geologically significant sink holes and caves off British Honduras and Andros, Bahama Islands, *Proceedings of the 3rd International Coral Reef Symposium*, Miami, 2, 238–242.

A.S. GOUDIE

BOLSON

Derived from the Spanish word for 'purse', bolsons are depressions with centripetal drainage that are surrounded by hills and mountains (Tight 1905). At their centre there is normally a saline playa or PAN, but if the low-lying area is drained by an ephemeral stream the basin may then be termed a 'semi-bolson' (Tolman 1909). Bolsons are a feature of semi-arid basin-and-range terrain and may contain such landform types as PEDIMENTS, ALLUVIAL FANS and BAJADAS.

References

Tight, W.G. (1905) Bolson plains of the southwest, *American Geologist* 36, 271–284.

Tolman, C.F. (1909) Erosion and deposition in the Southern Arizona bolson region, *Journal of Geology* 17, 136–163.

A.S. GOUDIE

BORING ORGANISM

Several life forms have evolved a means of penetrating a variety of substances for security and protection, wood and softer rocks being common subjects of such actions. The geomorphological interest is largely focused on the rock-boring organisms because their activity acts as a direct erosional agent and can also weaken the rock, making it more susceptible to erosion by other means. There exist terrestrial boring organisms, mainly algae and the fungal component of lichens, and particular interest has been shown in the marine borers, especially around the intertidal zone where they can lead to the formation of an undercut notch on rocky coasts. Some erosive mechanisms appear to be mechanical but many also appear to be chemical, attacking the more soluble rocks. Thus much of the interest in boring organisms lies in the field of the production of surface textures and smaller scale landforms in terrestrially exposed limestones (Trudgill 1985, Ch. 2, 3, 4, 8; Viles 1988) and in coastal limestone geomorphology where significant features, such as undercut notches of up to a few metres in dimension, can be formed (Trudgill 1985, Ch. 9, 10).

In studies involving environmental reconstruction, the occurrence of fossil intertidal boring organisms either above or below present sea level can provide evidence of former sea levels. This is particularly the case when undercut notches are found on dry land, considerably above present sea level. In some situations, these could have been formed by river action but where there are fossil perforations made by boring organisms, this confirms a marine origin for the undercut as assemblages of boring organisms in hard rock are unusual in fresh-water situations. This is especially the case if fossil boring bivalve shells are still present and the species can be identified and confirmed as marine organisms. In some cases, the shell material can be extracted and used for dating purposes and thus if there is a sequence of raised shorelines, palaeoenvironmental reconstruction is greatly assisted. Rowland and Hopkins (1971) noted that the boring bivalve mollusc *Hiatella arctica* can be found widely in the Arctic and Atlantic oceans and also in the Pacific Ocean from Alaska to Mexico. They noted the potential for the use of its fossil shells in paleoclimatic reconstruction.

Boring algae

Boring algae may be found in the first few millimetres of very many rock surfaces, and indeed the darker colour of rock surfaces which have been exposed for any length of time is often ascribable to this algal layer. The algae are frequently found to be blue-green algae (cyanobacteria). The algae associated with rock surfaces can be described as epiliths which live on the surface or as endoliths which live below it, with a further distinction being made for the chasmoliths which exist in the interstitial spaces between the rock grains – thus only endoliths which penetrate grains, or perforants, can be regarded as borers. The presence of endolithic perforant algae in limestone leads to the formation of very fretted surfaces known as phytokarst (Viles 1988). This can give a ferociously sharp, intricate and spongy rock surface. In cave entrances, the phytokarst is directional and angled to the light source and the borings are a product of erosion by phototrophic algae.

The benefit to the algae is to have access to moisture within the rock; however they still have to photosynthesize so they are found in thin layers just below and parallel to the surface at depths where moisture is present and where light can still penetrate. This optimum depth is termed the light compensation depth or LCD. The access to moisture is especially important in harsh environments and so while endolithic algae are found very widely over the rock surfaces of the Earth, they can occur in extreme environments including Antarctica (Friedmann and Ocampo 1976) where they can be important primary producers.

In the marine environment they commonly dominate in the mid- to upper shore, beyond which (inland) it is too dry and below which (towards the sea) it is wet enough for other organisms also to occur. They provide food for a wide range of rasping molluscs which contribute to rock erosion by ingesting rock with the algae. The algae then penetrate further into the rock to achieve an optimum LCD.

Boring endolithic algae are most usually found in carbonate rocks and the mechanism by which they bore is thought to be one where the algal filaments, about 10 μm wide, release extracellular chelating or acid fluids from the terminal cell. Using a high-powered electron microscope it can be established that up to 50 per cent of a rock surface bored by algae can be void space. Such boring can give rise to an extremely fretted dissected surface.

Boring fungi and lichens

The fungal portions of lichens can penetrate into rocks, again mainly in the carbonate rocks, in both intertidal and terrestrial environments. In both cases they exploit the weakness of calcite crystal interfaces and can also make larger pits.

Boring sponges

Boring sponges, commonly of the species *Cliona* are less able to withstand desiccation than boring algae and thus their distribution is from the mid- to lower intertidal. Extensions of the sponge tissue, termed etching amoebocytes, which, using acid secretions, are able to penetrate calcite in semicircular cuttings about 60–80 μm wide. On the surface, small 'keyhole' slots of 0.5–1 mm long are visible to the naked eye.

Boring bivalves and boring barnacles

Species of bivalve molluscs produce tubular borings which may penetrate into carbonate rock by several centimetres; they may also bore into live coral, sandstone, clay, peat and wood. There is evidence for acid secretion in carbonate substrates but they can also excavate the substrate by mechanical means, combining a rocking or rotational movement, which acts to grind the substrate, and a pumping motion using muscular contractions. In tropical areas the commonest boring bivalve is *Lithophaga* and the commonest boring barnacle is *Lithotrya*. In temperate regions the boring bivalve *Hiatella arctica* is a frequent borer of limestone in low intertidal and subtidal locations (Trudgill and Crabtree 1987). Additionally there also exist boring sipunculid worms and polychaete worms which generally make much thinner borings than the boring bivalve molluscs or boring barnacles.

Boring echinoderms

In the lower intertidal, subtidal and in rock pools several species of boring echinoderms exist, in temperate regions commonly *Paracentrotus lividus* (Trudgill *et al.* 1987) and in tropical regions *Echinometra lucunter* is common. The former make semicircular pits a few centimetres in diameter and the latter effect grooves in the rock surface. It is evident that echinoderms bore for protection, such as on exposed coasts of Carboniferous Limestone in Co. Clare, Eire, *Paracentrotus lividus* bores at rates between 0.25–1.5 $\mathrm{cm\,a^{-1}}$ whereas on sheltered coasts they may exist on unbored surfaces or in shallow depressions with lower excavation rates (Trudgill *et al.* 1987).

Rates of boring

The rates of erosion by boring organisms can be measured in a linear fashion ($\mathrm{mm\,a^{-1}}$) where there is surface retreat or a single boring, or in a cubic fashion ($\mathrm{cm^3\,a^{-1}}$) where there are more diffuse excavations. Published rates of the erosion of limestones by boring organisms (Spencer 1988) include the following:

Echinoderms	0.25–14.0 $\mathrm{cm^3\,a^{-1}}$
Sponges	1.0–1.4 $\mathrm{cm^3\,a^{-1}}$
Hiatella	5–10 $\mathrm{mm\,a^{-1}}$
Lithophaga	9–15 $\mathrm{mm\,a^{-1}}$
Lithotrya	8–9 $\mathrm{mm\,a^{-1}}$

Given that overall surface retreat ranges from around 0.5 to 4 $\mathrm{mm\,a^{-1}}$ and commonly is 1.0 $\mathrm{mm\,a^{-1}}$, it can be seen that boring organisms are highly significant erosive organisms. Indeed, where boring organisms are present, this leads to the formation of a horizontal undercut or notch in the coastline, not only through the direct action of the organisms themselves but also through the removal of the mechanically weakened rock by wave action. The mechanical boring of hard rock by the larger shelled organisms produces significant quantities of fine carbonate sediment.

Zonation of boring organisms

The zonation of intertidal organisms is of interest to the geomorphologist if it is proposed that there is a cause and effect relationship between biological zonation and morphological zonation (Trudgill 1987). Originally it was thought that biological zonation was a response solely to the ability of different species to withstand emersion and desiccation. However, theories of intertidal distribution have become less environmentally deterministic and now involve concepts of interspecific competition and predation. Thus species distribution can vary markedly in any one tidal zone according to the presence or absence of predators and other species. This suggests that while there may be a general zonation of boring organisms and hence morphological types produced by them, the distribution of individual boring organisms is liable to vary at different locations in relation to predation and competition rather than just to tidal zonation.

In addition, the variation of the landform itself may provide different micro-habitats which afford protection or sites which are too exposed, leading to local variability and that feedback effects can occur. In particular, on flatter, near-horizontal surfaces boring can produce low-lying areas which facilitate water retention and hence survival, meaning that they then become even deeper as boring activity and number of boring organisms can intensify.

Geomorphological significance

Schneider (1976) suggests that in the Adriatic, moisture conditions provide the limiting factor on boring activities. He sees three stages:

1. The primary depressions in rock surfaces are first colonized since they retain moisture longer than their higher surroundings. Conditions for boring and grazing prevail longest here. As a result each depression becomes the site of more intense biological erosion.
2. The pools enlarge laterally and small depressions coalesce, wet areas are preferentially bioeroded and relief is thus intensified.
3. This is the stage of maximal relief. It shows maximal contrast in ecological conditions and thus in the destructive processes. The water in the depressions is changed at most high tides and thus brings fresh sea-water to organisms. Each pool may enlarge and break into the next.

The sequence may be restarted if a deeper bedding plane or other weakness is reached, thus draining the pool; alternatively the rim of the pool may be breached.

Tidal range itself and the degree of exposure are also important considerations. In areas with limited tidal range, as, say, in the Mediterranean, there is commonly a deep undercut notch formed by boring organisms limited to 15–20 cm in height in the mid-intertidal which in itself may only have an amplitude of some 30 cm. The same ratio of notch to range applies as tidal ranges expand. Where coasts are exposed to larger waves and storms, boring organisms may be present but contribute quantitatively far less to the overall erosion of the coast, mechanical erosion tends to dominate, and, correspondingly, the undercut notch is either weakly developed or absent. Here the coast has the appearance of a sloping ramp rather than of a vertical cliff with recess or undercut notch as tends to be the case in sheltered locations.

References

Friedmann, E.I. and Ocampo, R. (1976) Endolithic blue-green algae in the dry valleys: primary producers in the Antarctic desert ecosystem, *Science* 193, 1,247–1,249.

Rowland, R.W. and Hopkins, D.M. (1971) Comments on the use of *Hiatella arctica* for determining Cenozoic sea temperatures, *Paleogeography, Paleoclimatology, Paleoecology* 19, 59–64.

Schneider, J. (1976) Biological and inorganic factors in the destruction of limestone coasts, *Contributions to Sedimentology* 6, 1–112.

Spencer, T. (1988) Coastal biogeomorphology, in H. Viles (ed.) *Biogeomorphology* 255–318, Oxford: Blackwell.

Trudgill, S.T. (1985) *Limestone Geomorphology*, Harlow: Longman.

Trudgill, S.T., Smart, P.L., Friederich, H. and Crabtree, R.W. (1987) Bioerosion of intertidal limestone, Co. Clare, Eire. 1: *Paracentrotus lividus, Marine Geology* 74, 85–98.

Trudgill, S.T. and Crabtree, R.W. (1987) Bioerosion of intertidal limestone, Co. Clare, Eire. 2: *Hiatella arctica, Marine Geology* 74, 99–109.

Trudgill, S.T. (1987) Bioerosion of intertidal limestone, Co. Clare, Eire. 3: zonation, process and form, *Marine Geology* 74, 111–121.

Viles, H.A. (1988) Organisms and karst geomorphology, in H. Viles (ed.) *Biogeomorphology*, 319–350 Oxford: Blackwell.

STEVE TRUDGILL

BORNHARDT

Bornhardts are dome-shaped, steep-sided hills, usually built of massive igneous rocks such as granite or ryolite, with bare convex slopes covered with very little talus and flattened summit surface. Bornhardts form due to differential weathering and erosion, in the course of which the surrounding less massive rock is eroded away leaving massive, sparsely jointed compartments. Many bornhardts have probably formed through selective DEEP WEATHERING followed by stripping of the SAPROLITE, but they can also emerge through gradual lowering of the surrounding terrain in the absence of deep weathering. Bornhardts are structure-controlled landforms and occur in every climatic zone; the existence of massive rock compartments is the necessary factor.

Characteristic features of bornhardts are slope-parallel joints called sheeting joints. They tend to be considered as resultant from unloading and would develop at shallow depth and after exposure, although it is also argued that sheeting develops in deeper parts of the crust, in response to compressional stress (Vidal Romani and Twidale 1999). Gradual opening of joints promotes slope instability, therefore rock slides and falls involving large masses of rock are common on bornhardts. Consequently, whereas upper slopes are bare and talus-free, footslopes may be covered by big blocks derived from upslope.

One of the persistent problems in the literature is the distinction between a bornhardt and an INSELBERG. The term 'bornhardt' was used by B. Willis in the 1930s to honour a German explorer from the turn of the nineteenth century, W. Bornhardt, who had introduced the name 'inselberg', but primarily to emphasize a special category of massive, dome-shaped inselbergs. The term subsequently evolved to describe monolithic domes regardless of their degree of isolation in the landscape. Therefore, although the terms are occasionally used as synonyms, these two categories of hills should not be confused. Nor is it justified to restrict bornhardts to granite lithology. Whereas 'inselberg' emphasizes isolation in space, bornhardts need to have distinctive domed shapes. Hence there is only partial overlap between the two and there exist bornhardts which are not inselbergs, and vice versa.

Classic examples of bornhardts include domes in Rio de Janeiro, Half Dome in Yosemite Valley and Ayers Rock in Australia. They are also abundant within African shields.

Reference

Vidal Romani, J.R. and Twidale, C.R. (1999) Sheet fractures, other stress forms and some engineering implications, *Geomorphology* 31, 13–27.

Further reading

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

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

Twidale, C.R. and Bourne, J.A. (1978) Bornhardts, *Zeitschrift für Geomorphologie N.F. Supplementband* 31, 111–137.

PIOTR MIGOŃ

BOULDER PAVEMENT

Striated boulder pavements form on intertidal surfaces affected by floating ice (Hansom 1983; Forbes and Taylor 1994), in ice-affected fluvial environments (Mackay and Mackay 1977) and at the base of glaciers or grounded ice sheets (Eyles 1988). Pavements deposited subglacially are the result of accretion of boulders around an obstacle and carry striations that are largely unidirectional, similar to fluvially derived striations produced by debris-charged floating ice. Striations on intertidal pavements are controlled by the direction of grounding of floating ice together with rotational striations imparted on stranding. Intertidal boulder pavements are composed of smoothed and highly polished boulders, often up to 1 m in diameter, tightly packed together as an undulating mosaic. They are often interrupted by bedrock outcrops together with furrows and polygonal depressions up to 5 m across. The main process seems to be the bulldozing and packing of loose boulders in the intertidal zone of a low-gradient boulder-strewn shore and the abrasion and striation of boulder surfaces by rock-shod floating ice. Prerequisites for their development appear to be a boulder source, frequent onshore movement of floating ice and a low-gradient intertidal zone. The degree of development is controlled by the frequency of onshore ice movement, well-formed pavements occurring in environments subject to high frequencies of freely moving ice.

References

Eyles, C.H. (1988) A model for striated boulder pavement formation on glaciated shallow-marine shelves: an example from the Yakataga Formation, Alaska, *Journal of Sedimentary Petrology* 58, 62–71.

Forbes, D.L. and Taylor, R.B. (1994) Ice in the shore zone and the geomorphology of cold coasts, *Progress in Physical Geography* 18, 59–89.

Hansom, J.D. (1983) Ice-formed intertidal boulder pavements in the sub-Antarctic, *Journal of Sedimentary Petrology* 53, 1035–1045.

Mackay, J.R. and Mackay, D.R. (1977) The stability of ice-push features, Mackenzie River, Canada, *Canadian Journal of Science* 14, 2,213–2,225.

JIM HANSOM

BOUNDARY LAYER

Emergence of boundary layer theory

The German engineer Ludwig Prandtl (1875–1953) presented a seminal paper to the 1904 Mathematical Congress in Heidelberg entitled 'Fluid Motion with very Small Shear' (Schlichting 1968). Prandtl showed that, with the aid of theoretical considerations and simple experiments, fluid flow over or around a solid body such as a sphere, cylinder or flat plate could be divided into two distinct regions. One region is relatively close to the body (or boundary), is relatively thin, and is characterized by large velocity gradients and viscous shear stresses. That is, fluid friction plays an important role in determining the physical characteristics of the layer. The second region is relatively far away from the boundary, and it is characterized by small velocity gradients and viscous shear stresses. That is, fluid friction may be neglected. This conceptualization termed boundary layer theory, as presented by Prandtl and further expanded by Geoffrey I. Taylor (1886–1975) and Theodor von Kármán (1881–1963), proved to become the foundation for modern fluid mechanics (Schlichting 1968).

A boundary layer can be defined as that part of the flow markedly affected by the presence of the boundary (Middleton and Wilcock 1994). Here, a flow refers to the motion of almost any kind of Newtonian fluid. Most real flows of geomorphic interest, such as flowing water in a river or blowing wind over a sand dune, are considered boundary layers because much of the flow is strongly affected by the boundary.

Laminar and turbulent flow

Osborne Reynolds (1842–1912) was the first to distinguish between two types of flow regime: laminar and turbulent (Schlichting 1968; Tritton 1988). In the laminar regime, the entire flow region appears to be divided into a series of fluid layers, each layer bounded by stream surfaces conforming to the boundary. In the case of two-dimensional flows, the traces of these surfaces on the flow plane are called streamlines (Figure 13). The rate of flow between two adjacent streamlines remains constant, although their spacing and orientation may vary, and velocity at-a-point does not vary or fluctuate in time. The transfer of fluid momentum, which results from the acceleration of slower fluid layers by faster moving layers, occurs at the molecular scale (Schlichting 1968).

In a turbulent flow, fluid particle paths are sinuous, intertwining and disordered. Fluid mixing occurs at both molecular and macroscopic scales. At this larger scale, fluid mixing commonly involves three-dimensional flow structures called eddies or vortices (turbulence), which are hairpin or horseshoe shaped rotating parcels of fluid moving away from or toward a boundary (Smith 1996). Because these vortices are relatively large and energetic, the time and length scales of turbulence are large and hence the turbulent transfer of fluid momentum is large compared to that of molecular diffusion. Within a turbulent flow, velocity at-a-point can fluctuate greatly as a result of passing vortices. Turbulence intensity typically is a measure of the magnitude of the velocity fluctuations compared to the time-average value at-a-point.

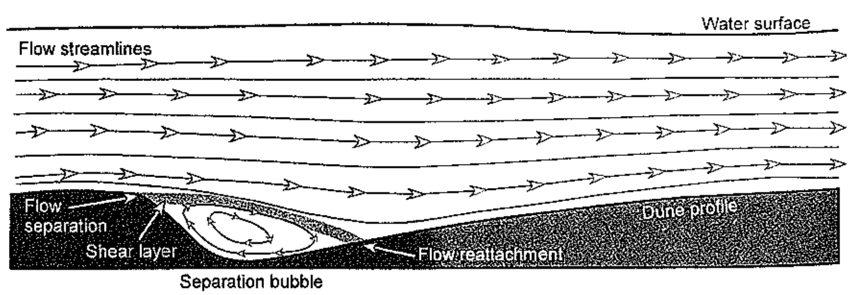

Figure 13 Time-averaged flow over a dune bedform (from Bennett and Best 1995). Flow is from left to right

The boundary Reynolds number

Reynolds found both experimentally and through dimensional analysis that the transition between laminar flow and turbulent flow occurs when the ratio of the inertia fluid forces is significantly larger than the viscous or frictional fluid forces (Tritton 1988). The inertia forces can be defined as the product ρud where ρ is fluid density, u is the mean flow velocity and d is the mean flow depth. The frictional forces can be characterized by the molecular viscosity of the flow, μ. This dimensionless ratio is called the boundary Reynolds number, Re. Flows are considered laminar when $Re < 500$, turbulent when $Re > 2000$, and transitional when $500 < Re < 2000$ (Tritton 1988). When a flow is laminar, any small disturbance to the flow, such as a protruding particle or a small change in bed topography, will not cause a change to the flow path lines or velocity and the disturbance will be damped by viscous forces. When a flow is turbulent, all disturbances to the flow will produce an effect throughout the boundary layer. When a flow is transitional, only select disturbances will affect the flow.

Few natural flows are laminar because most are deep and fast enough for the boundary Reynolds number to be very large. For example, the Mississippi River near its mouth has a Reynolds number near 10^7.

Flow separation and reattachment

As a flow moves along a curved boundary or over an obstacle, the boundary layer may separate and move away from the wall. Separation occurs when the pressure gradient in the downstream direction is adverse or unfavourable (Tritton 1988). A pressure gradient is considered adverse when a flow is expanding, diverging and decelerating in the longitudinal direction, and the pressure acting on the boundary is increasing, such as on the rear part of a streamlined pier. Both laminar and turbulent boundary layers can separate. Laminar flows usually require only a relatively short region of adverse pressure gradient to produce separation, whereas turbulent flows separate less readily.

Reattachment is the opposite of separation. There is a tendency for separation to be followed by reattachment unless the adverse pressure gradient continues long enough to prevent it (Tritton 1988). This separation–reattachment phenomenon is associated with a separation bubble or a region of flow recirculation. A typical example of flow separation can be found downstream of a ripple or a dune (Figure 13; Bennett and Best 1995). A recirculation bubble extends from the ripple or dune crest point to a distance downstream of about 5 to 7 times the bedform height, where flow reattaches to the boundary.

Downstream of the line of separation, there is a region of intense shear between the faster-moving outer part and the slower-moving or counter-rotating inner part of the boundary layer. Consequently, the flow along this shear layer is unstable, and turbulence is produced (Figure 13; Tritton 1988; Middleton and Wilcock 1994). Turbulent wakes, or regions of high turbulence, are typical of flow past any kind of obstruction at a high Reynolds number. The mixing layer present above a ripple or dune just downstream of the bedform brink is dominated by shear layer turbulence.

Structure of turbulent boundary layers

A turbulent boundary layer can be subdivided into three distinct zones: an inner layer, an outer layer and a wake region (Schlichting 1968). The inner region of a turbulent boundary layer is composed of a viscous sublayer (up to $y^+ = yu_*/\nu = 10$, where y is distance from the boundary, u_* is shear velocity, $u_* = (\tau/\rho)^{0.5}$, τ is bed shear stress, and ν is the kinematic viscosity of the flow) and a buffer layer (from $10 < y^+ < 40$). In the viscous sublayer, viscous forces dominate, yet very weak turbulent motions occur. These motions, called viscous sublayer streaks, are longitudinally oriented rotating tubes of alternating high- and low-speed fluid (Smith 1996). In some flows, sand streaks can be observed on the bed surface and these demarcate the location of low-speed streaks that tend to accumulate sand. Such sand streaks have been observed in the rock record and are called parting or current lineations.

The buffer layer is where the turbulent bursting process takes place. Low-speed streaks, in the general shape of a hairpin vortex, are lifted from the boundary and into the buffer region.

This lifted low-speed streak creates a thin shear layer that becomes unstable, oscillates, and is energetically ejected into the outer region of the flow (called a burst or an ejection event). Immediately following an ejection event, high-speed fluid from the outer region rushes in to replace the ejected fluid, impinging the bed (called a sweep event; see Smith 1996). This two-stage phenomenon is called the bursting process and can account for 70 per cent or more of all turbulence production within a boundary layer. Turbulent bursts or ejections are energetic enough to suspend sediment from the bed in river and airflows. Sweep events, with their high instantaneous drag forces, can entrain sediment particles resting on a bed surface.

In the outer region, representing the lower 10 to 20 per cent of the flow depth, there is a region where the velocity distribution varies logarithmically with distance from the bed, thus termed the logarithmic zone. Prandtl first conceptualized this velocity distribution in his 1925 mixing length theory (Schlichting 1968). Prandtl visualized a simple mechanism of fluid motion where parcels of fluid would move upwards or downwards, accelerating or decelerating the surrounding fluid. The distance over which the fluid is mixed is called the mixing length. Von Kármán expanded this theory by assuming that the mixing length varies as a simple function of distance from the bed multiplied by a dimensionless, universal coefficient (von Kármán's coefficient, $\kappa \sim 0.41$; Schlichting 1968). The final result is the Kármán–Prandtl law of the wall, $u/u_* = 1/\kappa \ln(y/y_0)$, where u is the velocity at a distance y from the wall and y_0 is the roughness height where velocity goes to zero. This velocity distribution has been shown applicable to a wide variety of flows such as pipes, rivers, near-shore environments, aeolian environments and in the near-bed region of gravity currents. Common uses of the law of the wall are the determinations of bed shear stress, roughness height and the turbulent mixing characteristics of the flow.

Finally, the velocity distribution in the outer 80 per cent of the turbulent boundary layer deviates from the logarithmic law. Here the velocity distribution is similar in shape to the velocity-defect profile in wakes (law of the wake; Coles 1956). For many straight rivers with relatively flat beds, the law of the wall is applicable over the entire flow depth (Middleton and Wilcock 1994). However, the presence of bedforms will significantly increase roughness length scales and velocity distributions.

Turbulent boundary layers are further qualified based on the roughness of the bed surface. This roughness parameter is called a grain Reynolds number Re_G, and it is defined as $Re_G = u_* k_s/\nu$, where k_s is the equivalent sand roughness height, which is approximately equal to the bed grain size (Schlichting 1968; Bridge and Bennett 1992). If the bed sediment is relatively small or absent, such that the grains are completely immersed in the viscous sublayer, then $Re_G < 11$, the sediment particles are subjected to viscous fluid forces only, and the boundary is considered hydraulically smooth. If the bed sediment is relatively large, such that the grains are larger than the viscous sublayer, then $Re_G > 70$, the sediment particles are subjected to turbulent fluid forces, and the boundary is considered hydraulically rough. Turbulent boundary layers are considered hydraulically transitional if some but not all grains are immersed in the viscous sublayer and $11 < Re_G < 70$. There are slightly different versions of the law of the wall and the determination of the equivalent sand roughness height depending on the roughness of the turbulent boundary layer. In general, beds that are composed of larger grains have relatively higher turbulent intensities and greater flow resistance.

The shape of the oft-used Shields curve for the dimensionless threshold of particle entrainment reflects this effect of grain Reynolds number on boundary layer characteristics (Bridge and Bennett 1992). Very small grains ($Re_G < 10$; hydraulically smooth flows) immersed in the viscous sublayer require higher dimensionless shear stresses for particle entrainment than larger grains that protrude higher in the turbulent boundary layer ($Re_G > 100$; hydraulically rough flows).

References

Bennett, S.J. and Best, J.L. (1995) Mean flow and turbulence structure over fixed, two-dimensional dunes: implications for sediment transport and bedform stability, *Sedimentology* 42, 491–513.

Bridge, J.S. and Bennett, S.J. (1992) A model for the entrainment and transport of sediment grains of mixed sizes, shapes and densities, *Water Resources Research* 28, 337–363.

Coles, D. (1956) The law of the wake in the turbulent boundary layer, *Journal of Fluid Mechanics* 1, 191–226.

Middleton, G.V. and Wilcock, P.R. (1994) *Mechanics in the Earth and Environmental Sciences*, Cambridge: Cambridge University Press.

Schlichting, H. (1968) *Boundary-Layer Theory*, 6th edition, New York: McGraw-Hill.

Smith, C.R. (1996) Coherent flow structures in smooth-wall turbulent boundary layers: facts, mechanisms and speculation, in P.J. Ashworth, S.J. Bennett, J.L. Best and S.J. McLelland (eds) *Coherent Flow Structures in Open Channels*, 1–39, Chichester: Wiley.

Tritton, D.J. (1988) *Physical Fluid Dynamics*, 2nd edition, Oxford: Oxford Science Publications.

SEAN J. BENNETT

BOUNDING SURFACE

Bounding surfaces represent discontinuities in sedimentation. Surfaces occur within all environments, and form an integral part of the geomorphic landscape and the rock record.

For example, migrating aeolian dunes produce three bedform-scale bounding surfaces. Reactivation surfaces form when the lee faces of dunes are eroded, such as when the dune reverses. Where associated with seasonal winds, these surfaces define cycles within the dune cross-strata. Superposition surfaces occur where smaller dunes migrate over the lee face of the main bedform. The surface is produced by scour associated with the passage of the INTERDUNE troughs of the superimposed dunes. Interdune surfaces begin with deflation of the stoss (windward) slopes of migrating dunes and culminate at the interdune floor.

At the dunefield scale, sequence or super surfaces form when accumulation in the field ceases. Examples include stabilization of the field by vegetation, and deflation to a planar surface defined by the water table.

Further reading

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

Rubin, D.M. (1987) *Cross-bedding, Bedforms, and Paleocurrents*, Tulsa: Society of Economic Paleontologists and Mineralogists.

SEE ALSO: interdune

GARY KOCUREK

BOWEN'S REACTION SERIES

In the early twentieth century, N.L. Bowen (1928) developed an idealized model, now called Bowen's Reaction Series, to describe the evolution or differentiation of igneous rocks. Recognizing that the types of minerals that form, and the sequence in which they crystallize, depend on the chemical composition of the magma and the temperature and pressure range over which the magma crystallizes, Bowen described two separate reaction sequences at high temperatures that eventually merge into a single series at cooler temperatures (see Figure 14).

The discontinuous series (left-hand side), involves the formation of chemically unique minerals at discrete temperature intervals from iron and magnesium-rich mafic magma. The first rocks to form are composed primarily of the mineral olivine. Continued temperature decreases, and fractionation of the magma (the early formed minerals are removed from the liquid by gravity), change the dominant minerals which form from pyroxene, to amphibole, and then to biotite.

The continuous series (right-hand side), involves the mineral plagioclase feldspar. At high temperatures, these minerals are dominated with calcium. With continued cooling, calcium and aluminum are exchanged for sodium and silicon. The convergence of both series occurs with a continued drop in magma temperature. Crystallizing rocks become richer in potassium and silica. The last mineral to crystallize in the Bowen's Reaction Series is quartz.

Examples of complete igneous sequences from basalt to granite are rare and other mechanisms are now known to produce differentiation sequences. Bowen himself acknowledged that the series was a simplification of very complex reactions and could be misleading if taken at face value. The reaction series also is used to explain susceptibility of minerals to weathering (see GOLDICH WEATHERING SERIES).

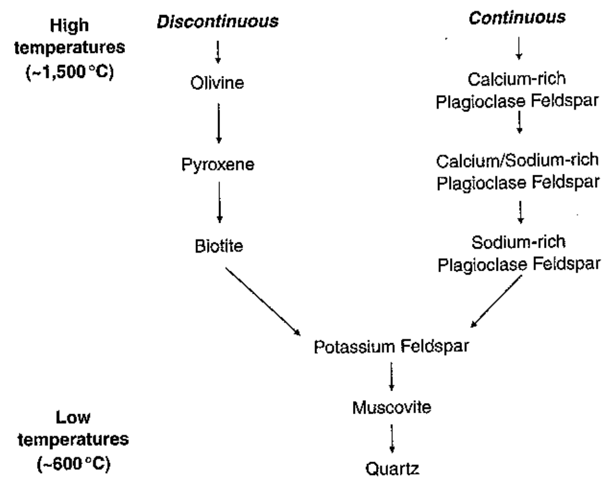

Figure 14 The Bowen's Reaction Series

Reference

Bowen, N.L. (1928) *The Evolution of the Igneous Rocks*, Princeton, NJ: Princeton University Press.

SEE ALSO: Goldich weathering series; chemical weathering

CATHERINE SOUCH

BOX VALLEY

A box valley has a broad flat floor, bounded by steep slopes which form a sharp piedmont angle. They are common in periglacial areas (see PERIGLACIAL GEOMORPHOLOGY), where they are formed by rapid lateral migration of braided channels, and by MECHANICAL WEATHERING by an 'ice rind' beneath the floodplain. Box valleys in mid-latitudes have been interpreted as relics of former cold climates. However, they also occur in tropical and arid lands, where they are formed by intense weathering or sheet wash.

Further reading

Büdel, J. (1982) *Climatic Geomorphology*, Princeton: Princeton University Press.
Young, R. (1987) Sandstone landforms of the tropical East Kimberley region, northwestern Australia, *Journal of Geology* 95, 205–218.

R.W. YOUNG

BRAIDED RIVER

Phenomenology of braiding

The hallmark of braided rivers is the presence of multiple active channels that divide and rejoin to form a pattern of gently curved channel segments separated by exposed bars (Plate 18). Braided rivers are marked equally by temporal dynamism: gradients in sediment flux associated with the complex spatial topography change local slopes, leading the flow to continually adjust its path as

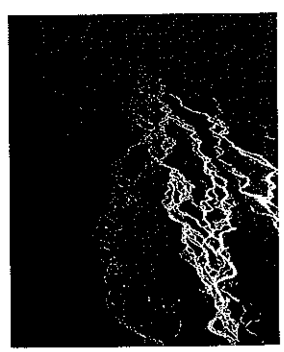

Plate 18 The braided Rakaia River, New Zealand

it picks its way through the network. Even when external conditions are constant, the braided pattern is continually changing, yet statistically consistent: a true dynamic equilibrium.

Braided rivers are known from around the world, but they are most common today at high latitudes. Often, braided rivers are GRAVEL-BED RIVERS, but prominent exceptions include some of the largest braided rivers in the world, such as braided sections of the Huang He and Ganges–Brahmaputra rivers. It has been suggested that braiding was the dominant river pattern on Earth before the first appearance of land plants in late Silurian time (Schumm 1968). Braided patterns have been observed on Mars, and vegetal patterns that resemble braiding are known from some bogs. However, although many morphological features of rivers are reproduced under oceans and lakes by the action of density currents, braiding appears to be rare or absent in the subaqueous realm.

It is worth distinguishing braided rivers from anastomosing channels, the other main type of anabranching channel (see ANABRANCHING AND ANASTOMOSING RIVER). In anastomosing channel networks, the typical width of the channels is much smaller than that of the bars, whereas in braided rivers these two length scales are comparable. Thus, the braided channel pattern is more space-filling than the anastomosed pattern. It is also worth distinguishing two somewhat different ways in which braided stream patterns can develop. In one case, 'confined braiding', there is a well-defined channelway that fills with water during floods and develops a pattern of submerged bars. As stage decreases, the bar tops emerge, producing a braided channel pattern. In 'free braiding', the braiding develops on an effectively unconfined plain. As discharge increases, an increasing number of channels is occupied, but the braid plain is never completely submerged. The relation between these two types of braiding is still not clear.

The dynamic character of braided networks owes much to interplay of their three basic elements (Ashmore 1991b): channel segments (anabranches), confluences and in-channel bars. Generally, these morphological elements are associated with locally parallel, converging and diverging bank geometries respectively. Channel segments may be straight or gently sinuous. Curved channel segments increase their SINUOSITY through erosion of their outer bank much as MEANDERING channels do, though braid anabranches often widen as they do so. Confluences are associated with channel narrowing, elevated velocities and local scour (Best 1988; Roy and Bergeron 1990). Bars are associated with channel expansion and widening, deposition (mainly on the bar periphery) and eventual splitting of the flow via scour along the bar sides. The dynamics of stream braiding largely results from the strongly nonlinear relation between flow strength and sediment flux. Confluences become scour sites because narrowing and acceleration of the flow increase its capacity for sediment transport. The scour further accentuates the narrowing and acceleration – an example of positive feedback. The converse is true in divergences. This tendency of the bed to accentuate local variability in the flow means the system never develops a static, steady-state configuration, even if water and sediment are supplied at a constant rate. This 'dynamic equilibrium' applies also to the flow of sediment through the braided network. As bars grow and are then incised by channels, sediment is impounded and released, producing highly variable sediment flow even if external conditions are steady (Ashmore 1991a).

Why do rivers braid?

Conditions commonly associated with the occurrence of braided rivers in nature include steep slopes, variable water discharge, coarse grain size and high rates of sediment supply. Empirically, we can identify sets of variables that discriminate braided from straight or meandering rivers. The most common of these discriminant plots is slope versus discharge, in which braided rivers appear at higher slopes for the same discharge than meandering rivers do. Empirical relations such as these provide hints as to the important variables, but little physical insight into the actual cause of braiding.

Historically, a major step in analysis of the causes of river patterns like braiding came with the application of stability analysis to the problem (Fredsoe 1978; Parker 1976). In stability analysis, one asks mathematically how a system responds to small perturbations. In analysing river planform the starting system is a straight channel, referred to as the 'base state'. Then one adds perturbations to the bed (and in some cases the banks), generally represented as one- or two-dimensional sine waves of infinitesimally small amplitude, and investigates how the perturbations change the flow and sediment-transport fields. If any of these perturbations changes the system in such a way as to produce its own growth, we have positive feedback and the system is unstable. This approach is based on the idea that natural systems are constantly being 'probed' by random disturbances – a tree falls in the river, for instance – that include a wide spectrum of wavelengths. A system that could not recover from such a disturbance would not last long in the real world.

In the case of rivers, the main control on planform stability turns out to be the channel aspect (width:depth) ratio. Channels narrower than about 20 times the depth tend to remain straight; those with widths roughly between 15 and 150 depths develop alternate bars, presumed to lead to meandering; and channels wider than about 150 depths develop multiple bars that are interpreted as leading to braiding.

Stability analysis was a major advance in that it provided a mechanistic foundation for understanding the origin of braiding and meandering. It also raised a number of new questions. For most rivers, the channel aspect ratio is not imposed from outside but is set by the dynamics of the channel itself. Unfortunately, the dynamics of channel width remains one of the fundamental unsolved problems of fluvial geomorphology. But it does seem clear that one of the strongest controls on width is the total effective sediment discharge (i.e. excluding the washload, suitably defined). Thus, high effective sediment loads are critical to braiding in two ways. First, high effective loads directly increase the width, directly increasing the aspect (width:depth) ratio. Second, for a given water supply, increasing the ratio of sediment to water discharge increases the slope, leading to smaller depths and thus further increasing the aspect ratio. This analysis helps explain why plots using slope and water discharge can discriminate a braiding 'regime' but suggests that neither variable is the fundamental control per se.

Chaos, complexity and braiding

The core idea of *chaos* in the scientific sense is that a fully deterministic system nonetheless can be effectively unpredictable. Surprisingly little has been done to analyse braided rivers formally as chaotic systems. It is clear that they are governed by a set of reasonably well-known deterministic equations, and they certainly appear to be unpredictable to any level of detail on timescales much longer than that required for migration of an anabranch or bar a significant fraction of its width. One especially fruitful line of analysis (Foufoula-Georgiou and Sapozhnikov 1998; Sapozhnikov and Foufoula-Georgiou 1996) has shown that braided-river plan patterns are fractal (specifically, *self-affine* fractals). The self-similarity or self-affinity that defines a pattern as fractal can occur either within one river (a small part of the river looks like a larger part), or between two different rivers (a small river looks like a larger river). An easily seen manifestation of similarity between large and small rivers is that the braided patterns one might see around town or on the beach share many basic dynamical characteristics with full-scale braided rivers. The similarity of large and small braided rivers makes braided rivers accessible to experimental study (Ashmore 1982).

Braided rivers also show a time–space scaling according to which the time evolution of a small part of a braided channel system is statistically indistinguishable from that of a larger part of the system, provided time is scaled (imagine speeding up or slowing down a film) according to a power of the ratio of the two areas being compared (Sapozhnikov and Foufoula-Georgiou 1997). These scaling results are not chaos per se, but power-law scaling of this kind is a common by-product of chaotic dynamics. The time–space scaling also implies that braided rivers may be self-organized critical systems.

Chaos theory arose from the study of atmospheric convection, and turbulent fluid flow remains one of the archetypes of chaotic behaviour. Braiding as a phenomenon seems analogous to turbulence in some respects (Paola 1996). In effect, a braided channel pattern is to a straight channel as turbulent flow is to laminar flow. Increasing the Reynolds number of laminar shear flow increases the momentum flux, which produces unstable high velocity gradients. The instability leads to a new, chaotic state that is more efficient at transferring momentum than the original laminar flow. In a straight river channel, increasing the sediment flux increases the width and (indirectly) decreases the depth, leading to unstable high channel aspect ratios. This instability leads to a new, chaotic state (braiding) that is more efficient at transferring sediment than a straight channel. (The nonlinearity of sediment flux as a function of flow velocity means that a flow system with high-speed and low-speed regions transports more sediment on average than a uniform stream with the same mean velocity.) The main stability parameter in braiding, and hence the equivalent of the Reynolds number, is the width:depth (aspect) ratio.

These observations help us understand aspects of the phenomenon of braiding not well captured in either empirical analyses or stability theory. The appearance of alternate bars in stability analyses is generally interpreted as implying a meandering plan pattern. But experimentally, meandering in channels without cohesive sediment is a transient phenomenon. Left to its own devices, a channel with alternate bars and noncohesive banks eventually evolves into a braided pattern with a low braid index. Channelized flow over noncohesive sediment cannot produce fully developed meandering, regardless of the channel aspect ratio. Evidently, just as pipe flow has two fundamental states (laminar and turbulent), channel flow in noncohesive sediment has two fundamental states: straight and braided. A fully realized meandering state requires that channels with alternate bars be stabilized, for example by cohesive sediment or vegetation.

We seem to be on the threshold of major advances in the theoretical modelling of braided rivers (see MODELS). The new field of complexity theory, which seeks unifying theoretical ideas and common behavioural patterns across a range of nonlinear systems, may help us to develop better theories of stream braiding. It is clear at this point that some aspects of the phenomenon of braiding can be captured in models that greatly simplify the detailed mechanics of flow and sediment transport (Murray and Paola 1994), while other aspects cannot. An insightful synthetic approach based on abstracting the detailed mechanics, perhaps ordered in the kind of hierarchical structure that has been used to study other complex systems, may be the most effective way of modelling braided rivers. It also may be that the best approach will simply be to develop numerical tools to solve the governing flow and sediment-flux equations on a sufficiently fine and adaptable mesh to allow for detailed simulation of the complex physics of braiding. Either way, newly emerging 'synoptic' field and laboratory data sets that capture the co-evolution of flow and topography over a whole river reach rather than a small area will be the standard against which new theoretical ideas will be tested. Of the main river types, braided rivers have proved to be the most challenging to analyse formally. The next edition of this encyclopedia will no doubt show dramatic results from some of the simulation efforts now under way.

References

Ashmore, P.E. (1982) Laboratory modelling of gravel braided stream morphology, *Earth Surface Processes and Landforms* 7, 201–225.

——(1991a) Channel morphology and bed load pulses in braided, gravel-bed streams, *Geografiska Annaler* 73A, 37–52.

——(1991b) How do gravel-bed rivers braid? *Canadian Journal of Sciences* 28, 326–341.

Best, J.L. (1988) Sediment transport and bed morphology at river channel confluences, *Sedimentology* 35, 481–498.

Foufoula-Georgiou, E. and Sapozhnikov, V.B. (1998) Anisotropic scaling in braided rivers: an integrated theoretical framework and results from application to an experimental river, *Water Resources Research* 34(4), 863–867.

Fredsoe, J. (1978) Meandering and braiding of rivers, *Journal of Fluid Mechanics* 84, 609–624.

Murray, A.B. and Paola, C. (1994) A cellular model of braided rivers, *Nature* 371, 54–57.

Paola, C. (1996) Incoherent structure: turbulence as a metaphor for stream braiding, in P.J. Ashworth, S.J. Bennett, J.L. Best and S.J. McLelland (eds) *Coherent Flow Structures in Open Channels* 705–723, Chichester: Wiley.

Parker, G. (1976) On the cause and characteristic scales of meandering and braiding in rivers, *Journal of Fluid Mechanics* 76, 457–480.

Roy, A.G. and Bergeron, N. (1990) Flow and particle paths at a natural river confluence with coarse bed material, *Geomorphology* 3, 99–112.

Sapozhnikov, V.B. and Foufoula-Georgiou, E. (1996) Self-affinity in braided rivers, *Water Resources Research* 32(5), 1,429–1,439.

Sapozhnikov, V.B. and Foufoula-Georgiou, E. (1997) Experimental evidence of dynamic scaling and indications of self-organized criticality in braided rivers, *Water Resources Research* 33, 1,983–1,991.

Schumm, S.A. (1968) Speculations concerning paleohydrologic control of terrestrial sedimentation, *Geological Society of America Bulletin* 79, 1,573–1,588.

CHRIS PAOLA

BROUSSE TIGRÉE

One of the most striking forms of PATTERNED GROUND is the brousse tigrée as identified from aerial photographs in West Africa (Clos-Arceduc 1956). This pattern is composed of alternating bands of vegetation and bare grounds aligned at the contour. From the air, these bands or arcs form a distinctive pattern similar to the pelt of a tiger.

Similar patterns have been recognized from aerial photos from many parts of the world. They were called *mulga groves* in Australia and *mogote* in Mexico. Ground truth may differ since banded vegetation can consist either of grass (Mauritania, Somalia, Sudan), shrubs (Australia, Mexico), or trees (Australia, Mali, Niger). They occur only where the co-occurrence of several critical conditions is met: low annual rainfall (75–650 mm), gentle and uniform slope (0.2–2 per cent) and crusting soils. These factors favour water runoff sufficient to produce sheet OVERLAND FLOW over a distance of a few tens of metres but insufficient to trigger the concentration of runoff into RILLS. In flatter landscapes, the vegetation is no longer banded but spotted because of the nondirectional runoff pattern. Slope also controls the wavelength (band plus interband width) of the pattern even at a local scale. The wavelength decreases exponentially with increasing slope gradient. Differences observed in the soils of bands and associated interbands are a consequence rather than a cause of banded ground (Bromley *et al.* 1997).

For a given slope, the mean annual rainfall determines the ratio between the width of the vegetation bands of arcs and the width of the bare bands. The bands accumulate runoff water and function as if they were in a higher rainfall climatic regime. The optimal rainfall for band development increases with increasing percentage of high rainfall event and decreasing duration of the rainy season. This optimal annual rainfall increases from 250 mm in central Australia to 550 mm in south-west Niger.

These banded patterns are natural examples demonstrating the principles of water, soil and nutrients conservation in space and time. Although the role of wind cannot be overlooked in certain circumstances, surface hydrological processes are critical to the ongoing functioning of banded landscapes. Three main processes are involved: differential infiltration, obstruction to overland flow, and efficient nutrient cycling. Soil crusts dominate in the interbands, resulting in low infiltration, whereas vegetation, litter and bioturbation effects facilitate high infiltration rates in the bands and arcs. The banded patterns act as a natural water harvesting system, the overland flow produced from the bare and impermeable interbands running onto the bands. Vegetation bands tend to obstruct or regulate sheet flow so that sediments and organic matter are continually being deposited and conserved within the bands, forming a natural bench structure that limits soil erosion. Due to the rainwater redistribution, the bands receive from two (in south-eastern Australia) to four times (in southwest Niger) the rainfall at the site. The centre of the bands has abundant biopores enabling effective water capture from the interband. The soils in the bands also concentrate more soil nutrients and organic matter than the adjacent interbands. This resource concentration enables the formation of a forest system, the productivity of which equals and can even double that of adjacent non-banded landscapes.

These systems can persist in the face of severe drought by adjusting the proportion of runoff and runon areas. They can also resist the stress and disturbance caused by moderate land use. The earliest indicator of deterioration is the decline in the contrast between the two mosaic phases. The late stage in degradation is characterized by disruption of the band pattern. Overgrazing is considered to be the prime cause of deterioration of banded landscapes in Australia. Firewood and timber harvesting threaten the brousse tigrée in West Africa.

Models have demonstrated that these patterns may result either from landscape degradation or rehabilitation, but the natural initiation of banded landscapes has never been observed. The slow upslope migration of the bands is also a debated topic. It is linked to the runoff/runon theory that underpins the basic functioning of banded vegetation. The obstruction of overland flow by the bands would favour the upslope germination of pioneer plants in this upslope edge and the decline of vegetation due to resource shortage at the downslope edge. This notion of upslope band migration is strongly supported by an array of arguments such as the seedling concentration on the upslope edge of the band, the decaying vegetation in the downslope edge, the sequence of soil crust types across the interbands, and the marked gradient in soil organic matter. The migration 'velocity' of bands has been assessed using a variety of methods including field monitoring with benchmarks, digitized aerial photographs, age distribution of trees with dendrochronology, and TRACERS (residual ^{137}Caesium) distribution in the soil, under a wide range of climatic and topographic conditions. The fastest observed migration was $1.5\,\mathrm{m\,yr^{-1}}$ for grass bands and $0.8\,\mathrm{m\,yr^{-1}}$ for trees and shrubs. Because of some stationary systems, the migration of vegetation bands cannot be regarded as an invariable property of the banded systems.

In the arid and semi-arid environment, the banded patterns are clear examples of heterogeneous landscapes that are more sustainable than homogeneous systems. The lessons drawn from them lead to the recognition of the ecological value of water harvesting and runoff farming.

References

Bromley, J., Brouwer, J., Barker, T., Gaze, S. and Valentin, C. (1997) The role of surface water redistribution in an area of patterned vegetation in South West Niger, *Journal of Hydrology* 198, 1–29.

Clos-Arceduc, M. (1956) Étude sur photographies aériennes d'une formation végétale sahélienne: la brousse tigrée, *Bulletin de l'IFAN*, série A 7(3), 677–684.

Further reading

Tongway, D.J. and Seghieri, J. (eds) (1999) *Acta Oecologica* 20(3), entire issue.

Tongway, D.J., Valentin, C. and Seghieri, J. (2001) *Banded Vegetation Patterning in Arid and Semiarid Environments. Ecological Processes and Consequences for Management*. Ecological Studies 149. New York: Springer.

Valentin, C. and Poesen, J. (eds) (1999) The significance of soil, water and landscape processes in banded vegetation patterning, *Catena* 37(1–2), entire issue.

SEE ALSO: crusting of soil

CHRISTIAN VALENTIN

BRUUN RULE

The prospect of accelerated sea-level rise as a consequence of global warming has renewed interest in models that link sea-level rise and coastal change, such as the shoreline translation model of Cowell and Thom (1994). But to date, no model is better known or more widely accepted than that of Per Bruun.

In 1962 Bruun (1962) proposed that with a rise in sea level, the profile of a beach and its nearshore zone would move landward and upward, and that the quantity of sediment eroded from the upper part of the profile would be transported seaward to build up the adjacent seafloor by an amount equivalent to the sea-level rise (Figure 15). In this model the retreat of the beach (R) is given as:

$$R = XS/Y,$$

where R is the difference in distance between the initial sea level–profile intercept and the intercept after sea-level rise, X is the horizontal length from shore to the limiting depth, S is the sea-level rise, and Y is the vertical dimension of the profile, which is the sum of the limiting depth below sea level and the top of the foredune above sea level.

Early testing of the model in a small-scale laboratory wave-table experiment, as well as sequential field measurements of shore profiles around Cape Cod, Chesapeake Bay and Lake Michigan during and after episodes of rising water level, tended to support the basic tenets of the model, such that in the 1970s it became known as the Bruun Rule, after temporarily being declared a 'theory' by Schwartz (1967). However, such status has not been without its critics, for rarely are the model's basic assumptions satisfied in the real world of multidimensional coastal morphodynamics (Healy 1991).

The Bruun Rule is a two-dimensional cross-shore model applicable to long straight sandy shorelines that adjust to a rise in sea level over decadal to centennial timescales. The original model assumes that the initial and final profiles are in equilibrium, that a constant profile shape is preserved over the period considered, that the total quantity of sediment in the cross-section is conserved, and that a constant water depth is maintained in the offshore zone as sea level rises. Clearly, these assumptions are unrealistic. For instance, given the fact that beach erosion is already widespread, it is unlikely that it would be possible to determine an EQUILIBRIUM SHORELINE, or that the shape of the profile would not change through time. It is also unrealistic to expect a beach profile to conserve sediment and

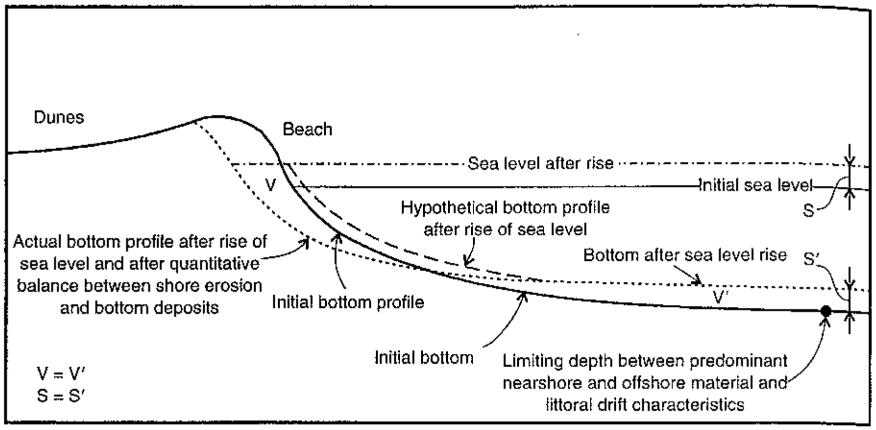

Figure 15 The Bruun Rule implies that the sediment volume removed from the beach and nearshore (V) must equal the sediment volume on the lower shoreface (V′) and that the lower shoreface aggrades in direct proportion (S′) to the rise in sea level (S). (After Bruun 1962 and Dubois 2002)

not to have sediment leakage, either as gains or losses resulting from LONGSHORE (LITTORAL) DRIFT, which is such a common process on sandy beaches. Similarly, wind erosion is not included in the Bruun Rule even though it can result in significant profile change through deflation of the beach and accumulation on a foredune at the landward end of a profile. Determination of the seaward end of a profile (closure depth) is equally problematical and maintenance of a constant water depth as implied in the Bruun Rule would result in a bathymetric or sediment discontinuity that in reality is difficult to define.

In spite of such difficulties, the Bruun Rule remains particularly attractive for several reasons. First, it is simple in concept and intuitively attractive given the fact that over the past hundred years or so global sea level has been rising at rates of around 1–2 mm per year, and that during that time approximately 70 per cent of the world's sandy shorelines have been eroding. Second, it can give quantitative results such that shore retreat will be 50 to 100 times the rise in sea level. For instance, a rise in mean sea level of 50 cm would result in beach recession of 25 to 50 m. And, third, the Bruun Rule has proven flexible enough to spawn a number of derivative models. Some refinements were proposed by Bruun (1983, 1988) himself, others by Dean and Maurmeyer (1983) who upscaled the concept to account for the landward and upward migration of an entire barrier island system. But the most persistent alternative model builder has been Dubois (1992, 2002).

Because the Bruun concept is not dependent on the shape of the shore profile, more complex topographies than in the original figure (Figure 15) can be incorporated (Dubois 1992). In Figure 16 zones of erosion and deposition are identified associated with onshore bar migration after shoreward displacement of the whole profile resulting from sea-level rise. Moreover, in this alternative model the eroded beach material is not only displaced seaward (as in the original Bruun model) but is also moved in a landward direction and washed or deflated on to the dune face or into a backing swale or lagoon. While overall conservation of sediment should be maintained within the whole profile, fine suspended sediment can be carried seaward and deposited in deeper water such that the lower shoreface and ramp do not accrete following the rise in sea level, but are simply abandoned by wave action (Figure 16).

In a comprehensive review of the response of beaches to sea-level changes, Working Group 89 of the Scientific Committee for Ocean Research (SCOR 1991) concluded that the quantitative predictions of shore change based on the Bruun

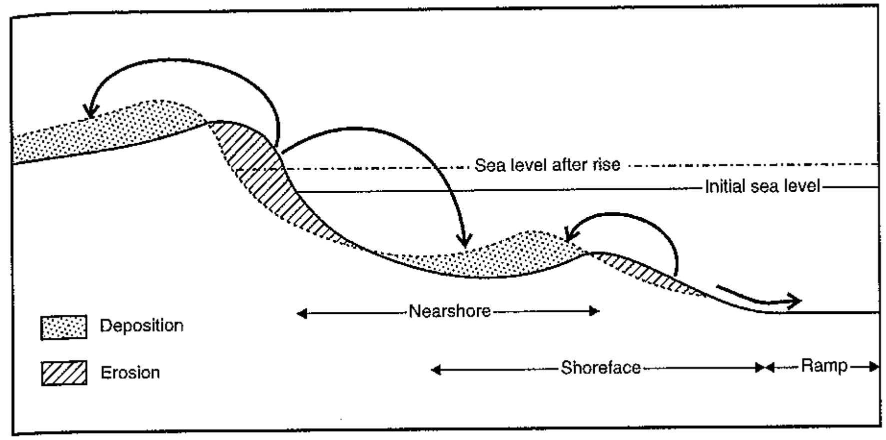

Figure 16 A two-dimensional model of a shore profile responding to a rise in sea level. Arrows show potential directions of sediment transport. Note that Bruun's Rule is embedded in this model, although it contributes only a small amount to the total shore erosion caused by rising sea level. (After Dubois 1992)

model are dependent on a number of parameters that are difficult to define, that there may be a significant time lag of beach response to sea-level rise, though the principal hindrance in achieving acceptable predictions is that the model does not include other sediment budget components that can result in either coastal accretion or erosion. Nevertheless, the SCOR group did suggest that the Bruun Rule could be used, though only for order-of-magnitude estimates of potential shore recession rates in appropriate coastal settings.

There is little doubt that globally sea level is rising and that sandy shores around the world are continuing to erode, including barrier beaches and barrier islands. The Bruun Rule gives an insight into how sea-level rise and coastal erosion are coupled, though we also know there are a host of other factors that contribute to coastal erosion independent of sea level.

References

Bruun, Per (1962) Sea-level rise as a cause of shore erosion, *Journal of the Waterways and Harbours Division*, Proceedings American Society of Civil Engineers 88, WW1, 117–130.

——(1983) A review of conditions for the use of the Bruun rule of erosion, *Coastal Engineering* 7, 77–89.

Bruun, Per (1988) The Bruun Rule of erosion by sea-level rise: a discussion of large-scale two- and three-dimensional usages, *Journal of Coastal Research* 4, 627–648.

Cowell, P.J. and Thom, B.G. (1994) Morphodynamics of coastal evolution, in R.W.G. Carter and C.D. Woodroffe (eds) *Coastal Evolution: Late Quaternary Shoreline Morphodynamics*, 33–86, Cambridge: Cambridge University Press.

Dean, R.G. and Maurmeyer, E.M. (1983) Models for beach profile response, in P.D. Komar (ed.) *Handbook of Coastal Processes and Erosion*, 151–166, Boca Raton, Florida: CRC Press.

Dubois, R.N. (1992) A re-evaluation of Bruun's rule and supporting evidence, *Journal of Coastal Research* 8, 618–628.

——(2002) How does a barrier shoreface respond to a sea-level rise? *Journal of Coastal Research* 18(2), iii–v.

Healy, T. (1991) Coastal erosion and sea level rise, *Zeitchrift für Geomorphologie* 81, 15–29.

Schwartz, M.L. (1967) The Bruun theory of sea-level rise as a cause of shore erosion, *Journal of Geology* 75, 76–92.

SCOR Working Group 89 (1991) The response of beaches to sea-level changes: a review of predictive models, *Journal of Coastal Research* 7, 895–921.

Further reading

Bird, Eric (2000) *Coastal Geomorphology: An Introduction*, Chichester: Wiley.

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

SEE ALSO: barrier and barrier island; beach–dune interaction; equilibrium shoreline; longshore (littoral) drift

ROGER F. McLEAN

BUBNOFF UNIT

Unit providing a useful means to quantify the rate of operation of diverse geomorphological processes as a rate of ground loss (perpendicular to the surface) or slope retreat. A unit equals 1 mm per 1,000 years, equivalent to $1\,m^3\,km^{-2}\,a^{-1}$ (Fischer 1969).

Reference

Fischer, A.G. (1969) Geological time–distance rates: the Bubnoff unit, *Geological Society of America Bulletin* 80, 594–652.

A.S. GOUDIE

BURIED VALLEY

A buried valley is the bedrock expression of a valley buried by more recent deposits. These features are surprisingly common but are not well known as they have no surface expression. They are usually identified following borehole information or other sub-surface investigations employing geophysical techniques. The identification of these features is often an important element in the reconstruction of the geomorphological history of an area.

A number of types of buried valley can be identified. First, there are those buried valleys which are the result of glaciation. These can be sub-aerially eroded valleys buried by deposits of glacial origin. A good example of this type is afforded in the English Midlands by the Proto-Soar valley. This broad sub-drift valley has its head between Stratford-on-Avon and Warwick and heads northeastwards towards Leicester where it underlies the contemporary river Soar, a major tributary of the Trent. Before the glacial event which buried this surface, the main watershed of England between Avon/Severn drainage to the west and Soar/Trent drainage to the east lay some 30 km at least to the west of its present position which is on top of a thick plug of glacial deposits. Elsewhere, buried valleys have been described which themselves have been created by subglacial processes and then subsequently been buried. In interpreting the buried valleys identified widely in East Anglia, Woodland (1970) drew attention to the 'tunnel valleys' of Denmark and northern Germany. In East Anglia, many of the buried valleys appear to be quite narrow, up to 500 m wide and often 100 m deep, whereas the features in Denmark are broader and shallower. Woodland, however, attributes a similar origin to these features – subglacial erosion beneath an ice sheet. In the case of the East Anglian examples, they have been infilled by a wide range of often complex sediments which mask the former topography. Other authors have preferred to explain the excavation of these tunnel valleys through glacial modification of existing valleys (West and Whiteman 1986).

Second, there are buried valleys resulting from changes in SEA LEVEL following the last glacial event. These are widespread around many coastlines where marine transgressions, estuarine deposition and alluvial fill have buried a landscape graded to a lower sea level. Thus under many existing rivers can be found buried valleys that represent the valley of the former river draining into a sea that might have been 100 m or so lower. Boreholes can reveal that the essentially level surface of the contemporary alluvium conceals an irregular surface comprising valley forms which often, but not always, parallel the existing drainage.

Finally, mention must be made of the numerous buried valleys, usually in urban areas, where human activity has been responsible for modification of the valley topography. In some cities, rubble has been used to infill minor valleys to create flat land for urban land uses.

References

West, R.G. and Whiteman, C.A. (1986) *The Nar Valley and North Norfolk*, Cambridge: Quaternary Research Association.

Woodland, A.W. (1970) The Buried Tunnel-Valleys of East Anglia, *Proceedings of the Yorkshire Geological Society* 37, 521–578.

TERRY DOUGLAS

BUTTE

Butte is a small steep-sided and flat-topped hill, built of flat-lying soft rocks capped by a more resistant layer of sedimentary rock, lava flow or duricrust, surrounded by a plain. Butte is smaller than MESA and may be considered as a more advanced stage of mesa degradation, although there are no formal criteria to distinguish between the two. Together with mesas, buttes are outliers, indicative of long-term scarp retreat. They occur in front of CUESTAs and plateau margins, their morphology being best pronounced in arid and semi-arid regions.

PIOTR MIGOŃ

C

CALANQUE

Coastal inlets (such as those to the east of Marseilles) which tend to be of a gorge-like form. They are widespread around the Mediterranean Sea and may be karstic dry valleys which have been partially drowned, as a result of the Flandrian transgression of the Holocene. Their positions may be fault controlled. In Mallorca they are called *calas*.

Further reading

Nicod, J. (1951) Le problème de la classification des 'calanques' parmi les formes de côtes de submersion, *Revue de Géomorphologie Dynamique* 2, 120–127.

Paskoff, R. and Sanlaville, P. (1978) Observations géomorphologiques sur les côtes de l'archipel Maltais, *Zeitschrift für Geomorphologie NF* 22, 310–328.

A.S. GOUDIE

CALCRETE

A term, proposed by Lamplugh (1902), to describe a terrestrial near-surface accumulation of predominantly calcium carbonate ($CaCO_3$) which occurs in a variety of forms ranging from powdery to nodular to highly indurated. It results from low temperature physico-chemical processes operating within the zone of WEATHERING which lead to the displacive and/or replacive introduction of $CaCO_3$ into a soil profile, sediment, rock or weathered material. Calcretes develop as a result of carbonates in solution moving laterally and vertically through vadose and shallow phreatic groundwater systems until they become, over time, saturated with respect to $CaCO_3$ and precipitate as calcite crystals (Wright and Tucker 1991). Calcretes often occur within soil profiles, where they may form single or multiple horizons, but they are not a type of soil. The term is synonymous with CALICHE (SODIUM NITRATE) and kunkur but distinct from other $CaCO_3$ cemented materials such as cave SPELEOTHEMS, lacustrine algal STROMATOLITES (STROMATOLITHS), TUFA AND TRAVERTINE, BEACH ROCK or AEOLIANITE.

Calcretes are estimated to underlie 13 per cent of the Earth's land surface and are most widespread in semi-arid regions. They form an important component of many contemporary dryland landscapes, and, where well indurated, may act as a threshold (see THRESHOLD, GEOMORPHIC) to erosion. Important areas of occurrence include the High Plains of the USA (e.g. Gile *et al.* 1966; Machette 1985), Africa north of the Sahara (e.g. Goudie 1973), the Kalahari of southern Africa (e.g. Watts 1980; Netterberg 1980), central and western Australia (e.g. Mann and Horwitz 1979; Milnes and Hutton 1983), and parts of southern Europe (e.g. Nash and Smith 1998). The close association between calcrete distribution and present-day dryland regions has led to the widespread use of calcretes in the geological record as indicators of past aridity. However, it is critical that the mode of origin of any calcrete is identified before it can be interpreted in this way. Whilst carbonate accumulation within soil may require a semi-arid climate, calcretes developed by other mechanisms may form under much wetter conditions. Non-pedogenic Holocene calcretes have, for example, been found in temperate locations such as the UK (Strong *et al.* 1992). Furthermore, calcrete accumulation is closely controlled by carbonate supply.

Calcretes are highly variable in appearance and range from thin rock coatings to massive horizons. Thickness varies with the mode of origin and stage of development, with laminar calcretes

rarely exceeding 0.25 m whilst multiple pedogenic and groundwater profiles may reach tens of metres thickness. Most calcretes are white, cream or grey in colour, though mottling and banding is common. Calcretes are predominantly cemented by calcite with some $CaMg(CO_3)_2$ (dolomite) often present. The size and shape of calcite crystals is dependent upon the composition of the host material, the duration of wetting and the influence of biological mechanisms. Cements are typically dominated by microcrystalline carbonate (or micrite) although larger crystals of sparry calcite may be present. If significant biological fixation of carbonate occurred during development, the calcrete is likely to exhibit a complex beta fabric dominated by organic structures when viewed in microscopic thin-section as opposed to simpler alpha fabrics developed by inorganic mechanisms (Wright and Tucker 1991). The mean global chemical composition of calcrete is *c*.78 per cent $CaCO_3$, 12 per cent SiO_2, 3 per cent MgO, 2 per cent Fe_2O_3 and 2 per cent Al_2O_3 (Goudie 1973), although variations occur dependent upon the host material chemistry, cement type, presence of authigenic silica and silicates, mode of origin, and stage of development.

There are a range of classification schemes for calcrete, of which the most widely employed use morphological criteria. Netterberg (1980), for example, recognized a range of forms including calcareous and calcified soils, powder, nodular, honeycomb, hardpan, laminar and boulder calcretes, a sequence which also reflects the phases of development of many calcretes. Gile *et al.* (1966) and Machette (1985) have proposed a scheme to assist the identification of calcretes at different stages of development, with stage I–III calcretes consisting of morphologically simple carbonate accumulations within soils progressing to more mature horizons by stages IV–VI. Calcretes have also been classified on the basis of their hydrological setting, with vadose, capillary fringe and

Table 5 A genetic classification of calcrete types

Environment of formation	Calcrete type	Incorporated calcrete types	Mode of formation
Pedogenic	Pedogenic calcrete	Caliche; kunkar; nari; petrocalcic horizons	Developed by vertical redistribution of calcium carbonate within a soil profile
Non-pedogenic	Non-pedogenic superficial calcrete	Laminar crusts; case hardening; gully bed cementation	Formed by surficial transport of calcium carbonate
Non-pedogenic	Non-pedogenic gravitational zone calcrete	Gravitational zone calcrete	Formed by downward accumulation of calcium carbonate in irregular permeability channels
Non-pedogenic	Non-pedogenic groundwater calcrete	Valley calcrete; channel calcrete; deltaic calcrete; lake margin calcrete; alluvial fan calcrete; fault trace and other groundwater calcretes	Formed by lateral transport of calcium carbonate
Non-pedogenic	Detrital and reconstituted calcrete	Recemented transported calcrete; calcretes which are brecciated and recemented *in situ*	Formed by recementation of existing fragmented or brecciated calcrete

Source: After Carlisle (1983)

phreatic types identified. Other schemes have subdivided calcretes by their dolomite content (Netterberg 1980) or by the relative abundance of alpha and beta cements (Wright and Tucker 1991). However, none of these classifications completely distinguishes between calcretes formed by different mechanisms. As such, the most helpful scheme is Carlisle's (1983) genetic classification (Table 5) which subdivides calcretes into pedogenic and non-pedogenic forms using geomorphological, chemical, macro- and micromorphological criteria.

Pedogenic calcretes develop near the land surface, usually in areas of low slope angle, through the mobilization, redistribution and relative accumulation of $CaCO_3$ within a soil profile. Formation may also involve some absolute accumulation of $CaCO_3$ if there are additional carbonate inputs to the profile. Such calcretes commonly show enrichment in $CaCO_3$ up-profile and consist of a powdery or nodular basal section overlain by a more massive hardpan which may, in turn, be capped by a laminar crust. Cements are usually dominated by micrite and, because of the mechanisms by which they develop, exhibit a complex micromorphology. Non-pedogenic calcretes encompass a wide variety of types, ranging from laminar crusts developed on rock or other calcrete surfaces by evaporation and/or biological fixing of $CaCO_3$, to detrital and reconstituted calcretes formed by the cementation of pre-existing fragmented crusts. By far the largest group are the groundwater calcretes. These are calcretes developed in channel, valley, alluvial fan, delta and lake marginal sediments, usually in the absence of soil-forming processes and sometimes at depths of tens of metres beneath the land surface. They can be distinguished from pedogenic calcretes by their lack of profile development, normally simple micromorphology and the presence of more crystalline calcite cements, especially where formation occurred at or below the water table (Nash and Smith 1998).

Despite the wide range of mechanisms by which they can develop, all calcretes result from the solution, movement and subsequent precipitation of $CaCO_3$, described by the following chemical reaction:

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

Calcretes require a carbonate source, usually released as a result of $CaCO_3$ dissolution (Goudie 1983). $CaCO_3$ solubility is closely linked to environmental pH, with solubility rapidly increasing below pH 9.0. Mechanisms which lower pH and drive the reaction to the right, such as the introduction of weak carbonic acid ($CO_2 + H_2O$ in the equation) or an increase in soil CO_2 partial pressure, will trigger dissolution. Sources of carbonate can be distant or local to the site of formation, and include weathered bedrock, volcanic (and other) dust and organic remains. Once carbonate is in solution, it may be moved laterally and/or vertically to the site of formation. Lateral transfer mechanisms may include transport in solution via ephemeral or perennial rivers as well as in shallow or deep groundwater systems. Vertical transfers include percolation of surface water or capillary rise from the water table. Carbonate precipitation (where the above reaction proceeds to the left) may be triggered by a variety of factors which lead to the concentration of carbonate-rich solutions and/or cause environmental pH to increase. Foremost amongst these are evapotranspiration, biological processes, decreases in CO_2 partial pressure, CO_2 degassing, and the common ion effect (Goudie 1983; Salomons and Mook 1986).

References

Carlisle, D. (1983) Concentration of uranium and vanadium in calcretes and gypcretes, in R.C.L. Wilson (ed.) *Residual Deposits: Surface Related Weathering Processes and Materials*, 185–195, London: Geological Society of London.

Gile, L.H., Peterson, F.F. and Grossman, R.B. (1966) Morphological and genetic sequences of carbonate accumulation in desert soils, *Soil Science* 101, 347–360.

Goudie, A.S. (1973) *Duricrusts in Tropical and Subtropical Landscapes*, Oxford: Clarendon Press.

——(1983) Calcrete, in A.S. Goudie and K. Pye (eds) *Chemical Sediments and Geomorphology*, 93–131, London: Academic Press.

Lamplugh, G.W. (1902) Calcrete, *Geological Magazine* 9, 575.

Machette, M.N. (1985) Calcic soils in the southwestern United States, in D.L. Weide (ed.) *Soils and Quaternary Geology of the Southwestern United States*, Geological Society of America Special Paper 203, 1–21, Boulder, CO: Geological Society of America.

Mann, A.W. and Horwitz, R.C. (1979) Groundwater calcrete deposits in Australia: some observations from Western Australia, *Journal of the Geological Society of Australia* 26, 293–303.

Milnes, A.R. and Hutton, J.T. (1983) Calcretes in Australia, in *Soils: An Australian Viewpoint*, 119–162, Melbourne: CSIRO/Academic Press.

Nash, D.J. and Smith, R.F. (1998) Multiple calcrete profiles in the Tabernas Basin, southeast Spain: their origins and geomorphic implications, *Earth Surface Processes and Landforms* 23, 1,009–1,029.

Netterberg, F. (1980) Geology of southern African calcretes: terminology, description, macrofeatures and classification, *Transactions of the Geological Society of South Africa* 83, 255–283.

Salomons, W. and Mook, W.G. (1986) Isotope geochemistry of carbonates in the weathering zone, in P. Fritz and J.Ch. Fontes (eds) *Handbook of Environmental Isotope Geochemistry*, Volume 2, 239–269, Amsterdam: Elsevier.

Strong, G.E., Giles, J.R.A. and Wright, V.P. (1992) A Holocene calcrete from North Yorkshire, England: implications for interpreting palaeoclimates using calcretes, *Sedimentology* 39, 333–347.

Watts, N.L. (1980) Quaternary pedogenic calcretes from the Kalahari (southern Africa): mineralogy, genesis and diagenesis, *Sedimentology* 27, 661–686.

Wright, V.P. and Tucker, M.E. (1991) Calcretes: an introduction, in V.P. Wright and M.E. Tucker (eds) *Calcretes*, 1–21, Oxford: Blackwell Scientific.

SEE ALSO: duricrust; silcrete

DAVID J. NASH

CALDERA

Calderas are large circular or elliptical volcanic depressions whose diameter (typically several or several tens of kilometres) greatly exceeds those of any included vents. They are formed by the evacuation of a magma chamber within the crust, and subsidence of the overlying rocks. Calderas can form on volcanoes of different magma composition, from low silica (mafic) to high silica (silicic), though the mechanisms vary. Though there are no hard and fast distinctions, the term crater tends to be reserved for smaller features created as a result of excavation of rock during explosive eruptions or smaller-scale collapse (collapse pits). The highest magnitude explosive eruptions on Earth, sometimes called super-eruptions, have generated the largest calderas. Volcanoes that have experienced more than one super-eruption are colloquially known as super-volcanoes. The geometries and structures of the resulting nested and overlapping calderas can be complex, and obscured by post-collapse uplift, volcanism and erosion.

Origins and development

Super-eruptions involve magma volumes of several thousand km^3 (masses up to 10^{16} kg). The rapid removal of such an amount of material from a crustal magma chamber invariably induces failure of the overlying rocks. There is a rough correspondence between the volume of magma erupted and that of the hole left in the ground by the caldera collapse. The largest known Quaternary eruption occurred about 74,000 years ago, and expelled an estimated 7×10^{15} kg (2,800 km^3) of silicic magma, making a significant contribution to the 100 km × 30 km caldera complex occupied today by Lake Toba in northern Sumatra (Oppenheimer 2002). Toba can certainly be classed as a super-volcano, since at least two similar events occurred around 840,000 and 500,000 years ago, as can Yellowstone (USA), whose last super-eruption took place about 600,000 years ago.

Important insights into caldera evolution associated with explosive eruptions have been gained from detailed investigations of a number of much smaller historic and prehistoric calderas, for example at Crater Lake (Oregon, USA; Bacon 1983), and Santorini (Greece (Plate 19); Druitt *et al.* 1999), and also ancient examples such as Scafell caldera of the English Lake District, Ordovician in age but revealing much of its structure thanks to erosion (Branney and Kokelaar 1994). Several subsidence processes have been recognized (Lipman 2000). Larger calderas tend to involve piston-like (plate) collapse, where the floor remains largely undeformed. The collapse occurs along steep ring faults, with vertical displacements of about 1 km. In contrast, downsag subsidence does not preserve the coherence of the developing caldera floor, which is instead tilted and flexed. Intermediate between these two is trapdoor subsidence, which occurs when the caldera floor remains hinged along part of its length but elsewhere has subsided in plate fashion. Geometrically complex systems of arcuate faults and subsiding blocks reflect, in some cases, the breakup of the floor during eruption but prior to ring-fault subsidence, and are referred to as piecemeal calderas (Branney and Kokelaar 1994). Relatively small-scale collapses resulting from modest explosive eruptions from a central vent, such as that of Pinatubo (Philippines) in 1991, are sometimes referred to as funnel calderas. These lack a bounding ring fault or coherent subsided plate.

Calderas associated with low silica (mafic) volcanoes such as those found in Hawai'i have somewhat different origins to their silicic counterparts. Some interpret their development as a late stage in the growth of Hawaiian shields; others see them as a recurrent process. Clearly, those on

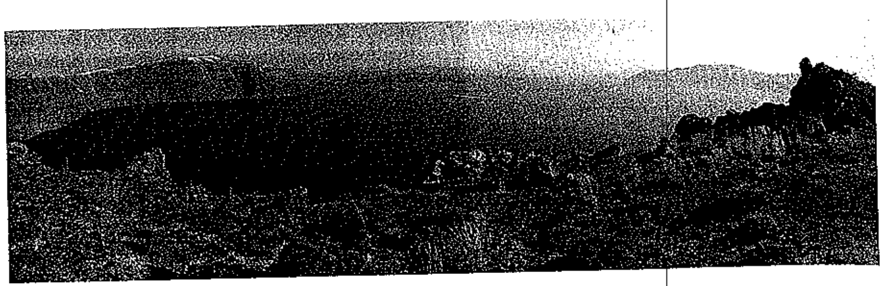

Plate 19 Santorini volcano, Greece. The last major caldera-forming eruption occurred in the mid-seventeenth century BC but is only the most recent in a series of eruptions exceeding 1,014 kg in magnitude (Druitt *et al.* 1999). The caldera rim is partly submerged such that the caldera is open to the sea

Mauna Loa and Kīlauea are very young features suggesting that they are rejuvenated at intervals of a few centuries, rather than the many millennia that can separate subsidence events on silicic volcanoes. Again, unlike silicic systems, the volumes of caldera subsidence on mafic volcanoes do not bear any obvious relationship with erupted volumes; the volumes of the young Hawaiian calderas are larger than any known Hawaiian eruptions. While Mauna Loa and Kīlauea appear superficially to be well-defined piston subsidence structures, Walker (1988) has suggested on the basis of mapping of the 2 Myr old Koolau volcano on Oahu, where erosion has exposed its 1 km depth roots, that downsagging may prevail in the central parts of funnel-shaped subsidence structures underlying the calderas. He suggests that, rather than simple piston subsidence into an evacuated magma chamber, the Hawaiian calderas develop as the dense intrusive rocks associated with the magma chambers sink into the warm lithosphere beneath.

Geomorphology

Calderas are enclosed by a topographic rim that is simply the head of the escarpment bounding the caldera. The inner walls can form cliffs in young calderas but retreat through time as a result of landslides. In map view, most large calderas reveal bites in the rim due to larger slope failures. The rock redistributed by landslides may form a collapse collar on the caldera floor. The arcuate bounding faults (ring faults) are sometimes exposed in deeply eroded calderas, and where observed, generally dip near-vertically or steeply inwards.

The largest calderas often enclose a central elevated massif. The vertical extent of the upheaval in these resurgent calderas can be 1 km or more. Toba is a good example; a lake occupies much of the caldera but an island – Samosir Island – occupies much of the lake. The island is composed substantially of the pyroclastic intracaldera fill from the Younger Toba Tuff eruption. Lacustrine sediments can be found several hundred metres above the present lake level and testify to substantial post-caldera uplift. Yellowstone is another example of a resurgent caldera. The mechanisms of resurgence, however, are not well understood. Refilling of the magma chamber is one possibility, along with the exsolution of remaining volatiles in the caldera eruption chamber and bubble formation (vesiculation) causing an increase in volume, expressed at the surface in the form of uplift (Marsh 1984).

Erosion rates of the pyroclastic deposits of caldera-forming eruptions can be rapid, especially for non-welded portions. Where developed, columnar jointing, akin to that observed in mafic lava plateaux (see LAVA LANDFORMS), can strongly influence the drainage fabric. In arid environments, the outflow sheets of ash flow deposits (i.e. those outside the caldera) may experience rapid aeolian erosion. This is evident in the wind-sculpted morphology of the Central Andean ignimbrites, which are adorned with YARDANGS and deflationary hollows. Wigwams or tent rocks are another common feature of the pyroclastic deposits of caldera-forming eruptions. These may develop by the intersection of drainage channels, or more commonly by the action of a resistant block of lava within the deposit, protecting the underlying material from erosion.

Hazards and climate change

Along with bolide impacts, large caldera-forming eruptions are the most catastrophic geologic events that affect the Earth's surface. Compared in terms of energy release, they are more frequent than bolides, however. A super-eruption today would have devastating impacts on regional populations and the global economy. The massive quantities of sulphur gases that can be released in super-eruptions strongly perturb atmospheric chemistry and radiation, with potentially global-scale climatic consequences (Rampino and Self 1993). Currently, little is known about the precursory events that might lead up to a super-eruption but numerous calderas worldwide have shown signs of unrest in the historic period (Newhall and Dzurisin 1988).

References

Bacon, C.R. (1983) Eruptive history of Mount Mazama and Crater Lake caldera, Cascade Range, U.S.A., *Journal of Volcanology and Geothermal Research* 18, 57–115.

Branney, M.J. and Kokelaar, P. (1994) Volcanotectonic faulting, soft-state deformation and rheomorphism of tuffs during development of a piecemeal caldera: English Lake District, *Geological Society of America Bulletin* 106, 507–530.

Druitt, T.H., Edwards, L., Mellors, R.M., Pyle, D.M., Sparks, R.S.J., Lanphere, M., Davies, M. and Barriero, B. (1999) *Santorini Volcano*, London: Geological Society.

Lipman, P.W. (2000) Calderas, in H. Sigurdsson, B.F. Houghton, S.R. McNutt, H. Rymer and J. Stix (eds) *Encyclopedia of Volcanoes*, 643–662, San Diego: Academic Press.

Marsh, B.D. (1984) On the mechanics of caldera resurgence, *Journal of Geophysical Research* 89, 8,245–8,251.

Newhall, C.G. and Dzurisin, D. (1988) *Historical Unrest at Large Calderas of the World*, US Geological Survey Bulletin 1,855, 2 volumes.

Oppenheimer, C. (2002) Limited global change due to largest known Quaternary eruption, Toba ≈74 kyr BP? *Quaternary Science Review* 21, 1,593–1,609.

Rampino, M.R. and Self, S. (1993) Climate-volcanism feedback and the Toba eruption of ~74,000 years ago, *Quaternary Research* 40, 269–280.

Walker, G.P.L. (1988) Three Hawaiian calderas: an origin through loading by shallow intrusions? *Journal of Geophysical Research* 93, 14,773–14,784.

Further reading

Francis, P. (1993) *Volcanoes: A Planetary Perspective*, Oxford: Oxford University Press.

Friedrich, W.L. (2000) *Fire in the Sea*, Cambridge: Cambridge University Press.

Long Valley Observatory http://lvo.wr.usgs.gov/

Williams, H. (1941) Calderas and their origin, *Bulletin of the Department of Geological Sciences, University of California* 25, 239–346.

Yellowstone Volcano Observatory http://volcanoes. usgs.gov/yvo/

SEE ALSO: lava landform; volcano

CLIVE OPPENHEIMER

CALICHE (SODIUM NITRATE)

The term has been used for both CALCRETE and for sodium nitrate deposits. The most famous and important deposits of the latter in the world occur in the Atacama Desert (Ericksen 1981), though others are known in California and Antarctica. Chile possesses a band of nitrate-containing terrain up to 30 km wide and 700 km long. Much of the Coastal Range is mantled with nitrate-bearing saline-cemented regolith, commonly ranging from a few tens of centimetres to a few metres in thickness. The petrography of the deposits, which cause extreme bedrock disintegration, is described by Searl and Rankin (1993). Most of the deposits lie at altitudes below 2,000 m, though some occur as high as 4,000 m. Low-grade deposits are also known in the coastal desert of Peru several hundred kilometres north of the Chilean border. Many of the deposits were extensively worked in the late nineteenth and early twentieth centuries. The landscape is now scarred with the pits from which nitrate was dug and is dotted with waste heaps.

The reason for the localization of the nitrate deposits appears to be the extreme aridity of the area, for sodium nitrate is more soluble in water than most common crust materials. The Atacama is among the driest and oldest of the world's deserts, and the average annual rainfall is less than 1 mm in the areas where the nitrate deposits are most prevalent. In any given part of the desert measurable rainfall (1 mm or more) may be as infrequent as once every 5–20 years. Heavy rainfall of a few centimetres or more may occur only a few times each century.

The Chilean nitrate fields occur typically in areas of low relief characterized by rounded hills and ridges and by broad, shallow debris-filled valleys. Significantly for their origin, they occur in all topographic positions from tops of hills and ridges to the centres of the broad valleys, though the richest deposits that have been worked most extensively tend to be on the lower slopes of hills.

Such a catholicity of geomorphological siting tends to imply that the nitrates have been derived as atmospheric inputs from the sea or from volcanic emissions, a mechanism supported by the fact that the nitrates occur on all rock and sediment types (Ericksen 1981). The model proposed by Ericksen (1981: 32) is that there has been long-term accumulation of atmospherically derived saline material for perhaps 10–15 million years (i.e. since the Mid-Miocene, under conditions of general extreme aridity). The sources of the material would include sea spray, volcanic emanations, photochemical reactions and dust from *salars* (salt lakes). The ore-grade nitrate deposits are formed by accumulation of saline materials on very old, flat to gently inclined or undulating landsurfaces, where rainwater dissolved the more soluble component and redeposited them at deeper soil levels. Ericksen's model of nitrate bedformation as a result of long-term atmospheric deposition has been confirmed by recent stable isotope studies (Böhlke *et al.* 1997).

References

Böhlke, J.K., Ericksen, G.E. and Revesz, K. (1997) Stable isotopes evidence for an atmospheric origin of desert nitrate deposits in northern Chile and southern California, USA, *Chemical Geology* 136, 135–152.

Ericksen, G.E. (1981) Geology and origin of the Chilean nitrate deposits, *United States Geological Survey Professional Paper* 1,188.

Searl, A. and Rankin, S. (1993) A preliminary petrographic study of the Chilean nitrates, *Geological Magazine* 130, 319–333.

A.S. GOUDIE

CALVING GLACIER

Calving GLACIERS terminate in water and lose mass by calving, the process whereby masses of ice break off to form ICEBERGS. Since they may be temperate or polar (see glaciers), grounded or floating, and may flow into the sea or into lakes, many types exist. They are widely distributed, but while lake-calving glaciers may exist in any glacierized mountain range, tidewater glaciers are currently confined to latitudes higher than 45°. Typically calving glaciers are fast flowing and characterized by extensional (stretching) flow near their termini, resulting in profuse crevassing. They terminate at near-vertical ice cliffs up to 80 m high. Calving activity above the waterline comprises a continuum from small fragments of ice to pillars the full height of the cliff. Below the waterline much ice may be lost through melting, but in deep water buoyancy causes infrequent but high magnitude calving events. In lakes, thermal erosion (melting) at the waterline can cause calving by undercutting the cliff. Calving permits much larger volumes of ice to be lost over a given time than melting.

Glaciers calve faster in deeper water. This correlation between calving rate (u_c in metres per annum) and water depth (h_w) is linear, and can be simply expressed as $u_c = ch_w$. The value of the coefficient c varies greatly in different settings, being highest for temperate glaciers and lowest for polar glaciers. Also, for any given water depth, calving is an order of magnitude faster in FJORDs than in lakes. The established correlation between calving rate and water depth may or may not imply that faster calving is *caused* by deeper water.

Calving glaciers are significant for three main reasons:

1 *Glacier dynamics* Calving glaciers comprise the most dynamic elements of many of the world's ice masses, and calving is the major means of ice loss from the two continental ICE SHEETs of Antarctica and Greenland. During the waning stages of Quaternary glacials (see ICE AGES (INTERGLACIALS, INTERSTADIALS AND STADIALS)), calving was the dominant process of mass loss around the mid-latitude ice sheets; the efficiency of calving helps to explain the catastrophic rates of ice sheet disintegration. Armadas of icebergs discharged during ice sheet collapses are believed to have caused global climate change by altering oceanic circulation.

2 *Non-climatic behaviour* Calving glacier fluctuations are highly sensitive to topographic controls (see PINNING POINT). Some tidewater glaciers fluctuate cyclically, in ways unrelated to climate, over distances of tens of kilometres and over timescales of centuries to millennia. Therefore neither the contemporary behaviour of calving glaciers nor the geomorphological records from past fluctuations (see MORAINE) are reliable indicators of climatic change.

3 *Socio-economic impacts* Calving glaciers constitute significant resources and GEOMORPHOLOGICAL HAZARDs for society. Resources include tourism and the potential for harnessing Antarctic tabular icebergs as a source of freshwater, while hazards include icebergs and OUTBURST FLOODs.

Further reading

Van der Ween, C.J. (2002) Calving glaciers, *Progress in Physical Geography* 26, 96–122.

Warren, C.R. (1992) Iceberg calving and the glacio-climatic record, *Progress in Physical Geography* 16, 253–282.

CHARLES WARREN

CAMBERING AND VALLEY BULGING

Cambering occurs where large-scale valley incision has exposed gently dipping bedrock in which competent strata overlie less competent strata. Cambered strata consist of an attenuating drape of competent caprock extending down the valley sides. This drape of CAPROCK shows evidence of extension, accommodated by deep fractures termed *gulls* (Figure 17). Gulls run parallel to the contours and separate intact caprock blocks. These blocks tend to tilt forward, increasing apparent dip and producing the widely reported 'dip-and-fault structure' (Figure 17). Cambering is often associated with the development of anticlinal deformation within the less competent strata beneath the valley axis, the resulting structure being termed *valley bulging*. Cambering and valley bulge structures are thought to have formed during the Quaternary Period. They reflect ice segregation processes within the less competent strata (usually clays) during PERMAFROST aggradation at depth, and subsequent thaw consolidation processes caused by thawing at the base of the permafrost (underthaw) during permafrost degradation in the transition from cold glacial to warm interglacial stages. Hutchinson (1991) has provided a detailed review of processes and structures involved in cambering and valley bulge development.

Classic descriptions of cambering and valley bulging come from the Jurassic Limestones and underlying Upper Lias clay in England (Chandler *et al.* 1976; Hollingworth *et al.* 1945; Horswill and Horton 1976). At the Empingham Reservoir, the cambered limestone strata are draped across much of the valley side. Gulls separate cambered blocks, which display classic dip-and-fault structure. Valley bulging is highlighted by disturbance of marker horizons within the Upper Lias clays. Disturbed brecciated clay extends to a depth of 25 to 30 m, the base lying parallel to the present ground surface, suggesting that the phase of brecciation occurred after the main valley relief was established. The disturbed Lias is underlain by a sheared plane of decollement at a depth of approximately 60 m under the valley crest and 30 m under the valley bottom. It is likely that the brecciated clay fabrics reflect the combined effects of ice segregation, creep and shear.

Hutchinson (1991: Figure 5.5) provides a model for cambering and valley bulging based on estimated displacements in the Lias at

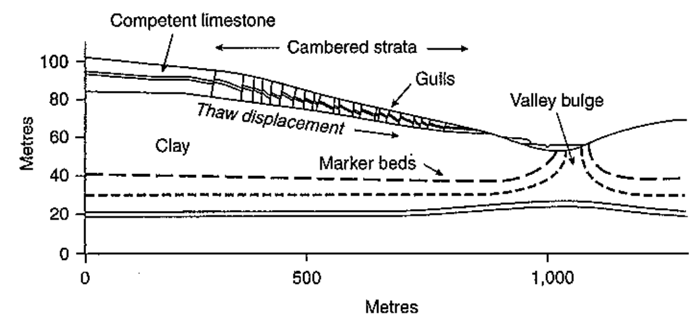

Figure 17 Cambering and valley bulging, based on the Gwash Valley, Lincolnshire, England (Horswill and Horton 1976)

Empingham (Vaughan 1976). The following stages are envisaged:

1 Initial incision of the river leading to unloading and valley rebound.
2 Permafrost develops, with ice segregation processes increasing ice contents of the frozen clay.
3 Valley-ward creep of the frozen clay occurs in response to lateral stresses. This causes extension and initial cambering of caprocks.
4 Permafrost degradation due to large-scale climate warming leads to thaw from the surface downwards, enhanced surface solifluction, and displacement of caprock down the valley sides over the thaw-softened clays beneath.
5 Thawing of the permafrost base due to geothermal heat flux is associated with thaw consolidation and high pore pressures within an effectively confined thawed stratum.
6 Lateral extrusion of the thawed clay at the permafrost base towards the valley bottom leads to compression along the valley axis and pronounced increase in the valley bulge structure.

Cambering and valley bulging represent some of the largest structures attributed to permafrost. Their presence as Quaternary relict features may be of considerable significance to engineering works such as the design of foundations for buildings which are particularly affected by the presence of gulls, and dam construction, where voids may increase water seepage, and deep-seated shearing may affect foundations.

References

Chandler, R.J., Kellaway, G.A., Skempton, A.W. and Wyatt, R.J. (1976) Valley slope sections in Jurassic strata near Bath, Somerset, *Philosophical Transactions of the Royal Society* A283, 527–556.

Hollingworth, S.E., Taylor, J.H. and Kellaway, G.A. (1945) Large-scale superficial structures in the Northamptonshire Ironstone Field, *Quarterly Journal of the Geological Society* 100, 1–35.

Horswill, P. and Horton, A. (1976) Cambering and valley bulging in the Gwash Valley at Empingham, Rutland *Philosophical Transactions of the Royal Society* A283, 451–461.

Hutchinson, J.N. (1991) Periglacial slope processes, in A. Forster, M.G. Culshaw, J.C. Cripps, J.A. Little and C.F. Moon (eds) *Quaternary Engineering Geology*, Geological Society Engineering Geology Special Publication 7, 283–331.

Vaughan, P.R. (1976) The deformations of the Empingham Valley slope, Appendix to P. Horswill and A. Horton (1976) Cambering and valley bulging in the Gwash Valley at Empingham, Rutland, *Philosophical Transactions of the Royal Society* A283, 451–461.

CHARLES HARRIS

CANYON

A canyon is a long, deep, relatively narrow, steep-sided valley, often cut through bedrock which forms precipitous cliffs along the valley walls. The word comes from the Spanish cañon. Canyons are formed by running water. The term is typically used for such features in arid and semi-arid regions, such as the western United States (e.g. the Grand Canyon in Arizona, USA). Canyons are similar to gorges (see GORGE AND RAVINE), but the side-walls are usually not as steep, and canyons are typically larger than gorges (e.g. the Grand Canyon contains an 'Inner Gorge' through which the Colorado River runs). Canyons are typical of mountainous regions, but are also found cutting high-elevation plateau (e.g. the Black Canyon of the Gunnison on the Colorado Plateau in Colorado, USA). They occur where stream erosion significantly outpaces weathering. Streams in canyons frequently flow through BEDROCK CHANNELS.

JUDY EHLEN

CAPROCK

Geomorphologically resistant lithological units, which protect underlying less resistant rocks from erosion and denudation, are called caprocks. Tablelands, CUESTAS, MESAS, buttes and hogbacks are examples of landforms which are composed of a backslope (dipslope), supported by a competent caprock, and a bipartite scarp slope. The scarp slope consists of an upper slope in the caprock and a moderately inclined lower slope in the less resistant rock below. It originates from fluvial downcutting or fault scarp development. In composite landforms like canyons and stepped cuesta landscapes whole sequences of caprocks and soft rocks can be found as in the Grand Canyon, in the Giant Staircase of southern Utah and northern Arizona or in the scarplands of southwestern Germany.

There is no specific method for defining caprock resistance on a metric scale, but there have been attempts to describe the resistance on an ordinal scale (Schmidt 1991). A scarp-forming rock must be relatively more resistant than the underlying soft rock. Caprock resistance can be connected with

lithological attributes such as: mechanical hardness protecting the rock against the direct effects of weathering and erosion; porosity and JOINTING resulting in greater water permeability. The attributes of the more resistant scarp-forming rock are most effective when the less resistant rock possesses contrasting characteristics such as mechanical weakness, easy disintegration and low permeability (Ahnert 1998: 239). The effects of different resistance are most visible in dry climates with selective weathering and erosion, where even minor lithological variations are reflected in slope geometry.

In most cases, especially in climates with greater surface water availability, permeability is more important for determining caprock resistance than mechanical strength. Due to their perviousness caprock outcrops are generally characterized by a lack of surface water and low values of drainage density. Water infiltrates and percolates through the caprock body until it reaches the impermeable lower slope rock. At the caprock/soft rock interface it reappears in springs and seepage zones, sometimes connected with sapping processes and slope undercutting. The missing to low activity of surface water erosion on the caprock-protected backslopes reduces mechanisms of denudational downwearing.

Most caprocks are sedimentary rocks like sandstones, conglomerates and limestones. The properties of the cementing materials (carbonates, iron oxides, clay minerals) control the erosion and weathering susceptibility of the sandstones and conglomerates. Karst processes are effective in carbonate caprocks. Joints and fissures are widened by solution resulting in increased permeability. In the scarplands the softer rocks below the caprocks are often clays, marls and fine, densely layered sandstones.

Volcanic rocks can also act as caprocks. This especially occurs in the case of lava flows which moved down former valley floors and covered older sedimentary rocks. The valley slopes in sedimentary material at the sides of the lava flows were subsequently removed by erosion and denudation, and the lava flows, due to their resistance, survived as caprocks of residual hills. This process is called relief inversion (see INVERTED RELIEF). Examples of this geomorphological process combination can be found in Tertiary volcanics in the Ore Mountains in Saxony and in Cenozoic volcanics at the margins of the Colorado Plateau in Utah and Arizona.

Scarp formation is also possible with DURICRUSTS as caprocks. The resistant crusts develop in homogeneous lithological units by weathering and soil-forming processes. Scarps of this kind are called homolithic scarps and are most frequently found in semi-arid areas (calcretes as caprocks) and in the tropics (silcretes and ferricretes). Some rock types can only have the function of caprocks under specific climatic conditions. Gypsum, for instance, can be a caprock in arid regions (e.g. southern Morocco), but in more humid climates rapid sulphate dissolution makes it an incompetent lithological unit.

The geomorphology of the upper scarp slope and the backslope is controlled by the attributes of the caprock. The upper scarp slope has a steep to cliff-like morphology. Its upper end forms a sharp crest (Trauf), especially in horizontally bedded caprocks with vertical joints. Caprock slopes, which are controlled by mass movement activity, also have a sharp crest at their top. In poorly cemented sandstones (e.g. Navajo Sandstone on the Colorado Plateau) rounded upper slopes are developed, which can also be found in more humid and colder areas, where sheet wash or solifluction processes are active on the upper slope.

The backslope begins at the scarp crest and follows the direction of caprock dip. In dry regions with highly selective weathering and erosion the inclination of the backslope is the same as the dip angle. Here the backslopes are stripped surfaces. The overlying less resistant rocks have been removed by erosion and denudation; there is a close conformity of topographic and structural relief. In the scarplands of more humid climates (e.g. Central Europe) the inclination of the backslope is generally less than the dip. The overlying strata are truncated at a low angle in the distal parts of the backslope. It resembles an inclined planation surface, but it is a caprock-controlled structural relief element of the cuesta landscape.

Especially in dry climates there is no mechanism of denudational downwearing of the caprocks. Their erosion is accomplished by aquatic and gravitational processes, which work on the scarp slope and result in lateral backwearing by parallel scarp recession. The area of the caprock outcrop is consumed by scarp retreat, at the margins of residual outliers by circumdenudation. The rate of retreat is controlled by caprock lithology and thickness (Schmidt 1989). It is not surprising that the backslope length is also controlled by these variables, and additionally by structural dip and the lithology of the overlying rocks (Schmidt 1991).

References

Ahnert, F. (1998) *Introduction to Geomorphology*, London: Arnold.
Schmidt, K.-H. (1989) The significance of scarp retreat for Cenozoic landform evolution on the Colorado Plateau, USA, *Earth Surface Processes and Landforms* 14, 93–105.
—— (1991) Lithological differentiation of structural landforms on the Colorado Plateau, USA, *Zeitschrift für Geomorphologie Supplementband* 82, 153–161.

Further reading

Blume, H. (1971) *Probleme der Schichtstufenlandschaft*, Darmstadt: Wissenschaftliche Buchgesellschaft.
Schmidt, K.-H. (1988) Die Reliefentwicklung des Colorado Plateaus, *Berliner Geograpische Abhandlungen* 49, 1–183.

SEE ALSO: butte; cuesta; hogback; mesa; structural landform

KARL-HEINZ SCHMIDT

CASE HARDENING

Case hardening describes rocks with outer shells more resistant to erosion than interior material. This hardening is sometimes called induration. Although case hardening is occasionally used as a synonym for DURICRUST, case hardening most frequently refers to differential weathering of the same rock type – often associated with intricate weathering features such as tafoni (Campbell 1999).

Two general types of processes create the appearance of case hardening: core softening of the interior; and case hardening of the exterior. James Conca proposed a lithologically based explanation. Crystalline rocks such as granite tend to core soften, whereas clastic rocks such as sandstone tend to case harden. The dichotomy has to do with the way the minerals bond together. Since sandstone grains are held together by cementing agents, a greater accumulation of cements at the surface causes case hardening. In contrast, the greatest change in hardness in a crystalline rock takes place when bonds are broken by CHEMICAL WEATHERING. Core-softened boulders are seen in locales of most intense chemical weathering.

The early literature advocates the view that hardening occurs by solutions that are mobilized from the rock's WEATHERING rind, drawn out by evaporative stresses, and then reprecipitated in the rock's outer shell. A growing body of evidence indicates that external agents also penetrate into the outer shell of the host rock, hardening the surface. A variety of different hardening agents have been found within host rocks lacking these agents. Amorphous silica, calcite, calcium borate, clay minerals such as kaolinite, iron hydroxides, oxalate minerals, rock varnish and other internally and externally derived agents penetrate about a millimetre to harden the very surface of the rock.

Case hardening, by definition, is not a ROCK COATING. However, a wide variety of rock coatings can act as case-hardening agents. Glazes of mostly silica and aluminum with some iron, only 20–30 μm thick, impede erosion of greenschist in southern England (Mottershead and Pye 1994). The role of silica glaze can be striking for temperate sandstones: '[o]ne of the most important characteristics of many porous sandstones is their tendency to case-harden owing to the development of a surface crust or rind' (Robinson and Williams 1994: 382). In Antarctica, iron-stained silica glaze reduces permeability and channels moisture towards uncoated rock surfaces. Thus, rock weathering is concentrated away from the rock coatings (Conca and Astor 1987). Lichen-generated oxalates protect sandstone surfaces in the Roman Theatre of Petra. Dark coatings of silica, oxides of iron/manganese, and charcoal case hardened rock faces at Yarwondutta Rock, Australia (Twidale 1982). While lichens are usually erosional agents, these epilithic (rock surface) organisms sometimes protect the underlying rock from erosion.

Although case hardening is most commonly noted in warm deserts where little soil covers rock surfaces, case-hardened rocks occur in all terrestrial weathering environments. In the wet tropics, for example, case hardening is frequently seen on bedrock along rivers at stages only reached by wet-season floods. In alpine settings, case hardening helps preserve glacial polish. Silica glaze is an important case-hardening agent in temperate (Mottershead and Pye 1994; Robinson and Williams 1994) and Antarctic areas (Conca and Astor 1987). Iron films can be seen splitting apart and also holding together weathering rinds in northern Scandinavia (Dixon *et al.* 2002).

Case hardening (Plate 20) often has subsurface origins in JOINTING. Mottershead and Pye (1994), for example, discerned a three-stage process. First, the host rock hardens along joint faces within the subsurface. Silica, aluminum and some iron comprise the bulk of the case-hardening agent. Second, DENUDATION of the land surface exposes joint faces at the surface. Third, erosion of rock underneath the case-hardened surface creates cavities called TAFONI, that highlight the case hardening. Rock engravings (petroglyphs) also emphasize planar JOINTING surfaces that were case hardened while in the subsurface. Road cuts of granitic rocks that weather to GRUS often reveal case-hardened subsurface joints. Geothermal and other DIASTROPHISM processes often leave behind case-hardened joints.

Considerable disagreement exists over how long it takes case hardening to form, with assertions in the literature ranging from months to thousands of years. James Conca studied rates of hardening in the Mono Basin of eastern California, finding that changes take place on the timescale of thousands to tens of thousands of years. Rates of hardening, however, vary with climate and the particular hardening process.

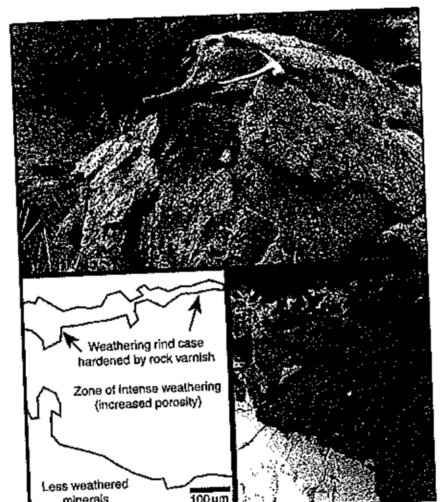

Plate 20 Case hardening on a *c.*140,000-year-old moraine boulder of the Sierra Nevada, California, where a combination of processes produce the differential weathering seen in the top image. Core softening of the host granodiorite boulder is the most important process. The electron microscope image and corresponding map shows a close-up of the area around the tip of the rock hammer. Some softening comes from chemical weathering and some hardening takes place as a result of the penetration of desert varnish into the weathering rind

References

Campbell, S.W. (1999) Chemical weathering associated with tafoni at Papago Park, Central Arizona, *Earth Surface Processes and Landforms* 24, 271–278.
Conca, J.L. and Astor, A.M. (1987) Capillary moisture flow and the origin of cavernous weathering in dolerites of Bull Pass Antarctica, *Geology* 15, 151–154.
Dixon, J.C., Thorn, C.E., Darmody, R.G. and Campbell, S.W. (2002) Weathering rinds and rock coatings from an Arctic alpine environment, northern Scandinavia, *Geological Society of America Bulletin* 114, 226–238.
Mottershead, D.N. and Pye, K. (1994) Tafoni on coastal slopes, South Devon, U.K., *Earth Surface Processes and Landforms* 19, 543–563.
Robinson, D.A. and Williams, R.B.G. (1994) Sandstone weathering and landforms in Britain and Europe, in D.A. Robinson and R.B.G. Williams (eds) *Rock Weathering and Landform Evolution*, 371–391, Chichester: Wiley.
Twidale, C.R. (1982) *Granite Landforms*, Amsterdam: Elsevier.

Further reading

Conca, J.L. and Rossman, G.R. (1982) Case-hardening of sandstone, *Geology* 10, 520–523.
Paradise, T.R. (1995) Sandstone weathering thresholds in Petra, Jordan, *Physical Geography* 16, 205–222.

SEE ALSO: chemical weathering; denudation; grus; jointing; rock coating; tafoni; weathering

RONALD I. DORN

CATACLASIS

A natural process whereby a faulted rock is deformed as a result of mechanical forces in the crust, such as fracturing, shearing, grain boundary sliding, and granulation. Cataclasis transforms a simple fault into a zone of fracturing and deformation without chemical alteration, and causes a decrease in the porosity of the rock alongside rock volume. Cataclasis takes place at low temperature–low pressure conditions, and high strain rates. The product of cataclasis in sediment is a cataclasite, a metamorphic rock composed of angular fragments (e.g. tectonic breccia) and a structureless rock powder fabric. When considered on a regional scale, cataclasis has also been interpreted as a flow mechanism.

STEVE WARD

CATACLINAL

A dip stream or a valley that runs in the same direction as the dip of the surrounding rock strata. Cataclinal slopes can be further classified into overdip slopes, underdip slopes, and dip slopes (steeper than, shallower than, and following the dip of surrounding strata, respectively). They may follow an individual rock layer from the base of a mountain to its peak (e.g. Mount Rundle, Canadian Rockies) (Cruden 2000). In contrast, anaclinal slopes dip in the opposite direction to the surrounding strata.

Reference

Cruden, D.M. (2000) Some forms of mountain peaks in the Canadian Rockies controlled by their rock structure, *Quaternary International* 68–71, 59–65.

STEVE WARD

CATASTROPHISM

Catastrophes in common parlance are unexpected major events with negative outcomes affecting large numbers of people. Large earthquakes, floods, hurricanes and wars are good examples of catastrophes, not all of which have geomorphic relevance. In the slightly more technical language of natural hazards, catastrophes are high magnitude, low frequency geophysical events which impact the socio-economic environment negatively. At a third level of technicality, catastrophes can be defined in the context of a mathematical approach to the analysis of inherently unstable Earth surface systems. This mathematical approach is called bifurcation theory and a special branch of bifurcation theory has been called catastrophe theory (Thom 1975). The essential concept is that many Earth systems are inherently unstable and are called dissipative structures. Such systems are characterized by a series of thresholds, below which the solutions to equations governing system dynamics are unique; beyond the threshold (or bifurcation point) the system loses its structural stability and undergoes a sudden or catastrophic change to a new form. Examples in the geomorphic literature include stream junctions in the Henry Mountains, Utah (Graf 1979), sediment transport processes in rivers (Thornes 1983) and accretion–degradation processes on a beach (Chappell 1978).

Catastrophism is a mode of thought which was common in the eighteenth and early nineteenth centuries and had as its basic premise that Earth history consisted of a series of high magnitude events separated by periods of quiescence. This mode of thought contrasts with gradualism in which small-scale, commonly acting processes are thought to be the dominant mode of geomorphic evolution.

The origins of catastrophism have often been traced to Baron Georges Cuvier (1769–1832) but, as we shall show below, his was only one brand of catastrophism and there were many catastrophist predecessors. Cuvier was the father of comparative anatomy and proposed that the Earth had suffered many catastrophes in the form of global earthquakes, each of which had changed the global landscape and annihilated almost all the flora and fauna. After each catastrophe, a new set of flora and fauna appeared: 'Thus we shall seek in vain among the various forces which still operate on the surface of our Earth, for causes competent to the production of those revolutions and catastrophes of which its external crust exhibits so many traces' (Cuvier 1817: 36–37). Huggett (1990) summarizes the issue by saying that there is either abrupt and violent change (catastrophism) or gradual and gentle change (gradualism). But the issue is more complex. There are different styles of catastrophism, and the style is determined by the premises of the author with respect to the following dichotomies: actualism v. non-actualism; directionalism v. steady state and, in dealing with organic change, internalism v. externalism. Although dichotomies tend to oversimplify the situation and in reality theorists of the Earth occupy positions along a spectrum of ideas, some simplification is necessary within the word constraints of this encyclopedia entry. The issue between actualists and non-actualists is whether past processes have differed in kind from those now in operation; non-actualists took the view that past processes could differ in kind from those presently in operation. The issue between directionalists and steady statists has been eloquently discussed by Gould (1987) in terms of time's arrow versus time's cycle; directionalists took the view that monotonic change, whether progress or regress, could be detected in Earth history. The issue between internalists and externalists, which arose only in the context of organic change, was whether the motor of organic change was an inner drive or external environmental factors; externalists argued that the environment forced change. Using these dichotomies as a basis for classification of styles of catastrophism, Huggett (1990) came up with eight categories of catastrophism which could be distinguished in the early history of environmental science. Six of these categories are reproduced here:

1. *Actualistic directional catastrophism* The Wernerian system of Earth history, following Abraham Werner (1749–1817), is a classic example of this kind of catastrophism. He envisaged five periods of Earth history, punctuated by intermittent and catastrophic ocean subsidence and precipitation of crustal rocks. The five periods were consecutive and demonstrated directional change.
2. *Non-actualistic directional catastrophism* René Descartes (1596–1650) described the origin of the Earth as an incandescent ball, followed by collapse of the Earth's outer crust and the release of massive volumes of water. His system involved directional evolution from original chaos created out of nothing by God through to an ordered universe which evolved according to natural laws invested in the original particles by God. The so-called Scriptural geologists, such as William Buckland (for most of his active career), Adam Sedgwick (1785–1873), William Conybeare (1787–1857) and Robert Murchison (1792–1871) all fell into this category in the sense that they believed in God's special intervention in the regular course of Nature through geological catastrophes and the sudden rise of species.
3. *Non-actualistic steady-state catastrophism* Baron Georges Cuvier (1769–1832) was the leading protagonist of this school of thought. He saw in the fossil and stratigraphic record evidence of catastrophic changes too great to be explained by the ordinary, slow-acting processes on the Earth's surface. Each catastrophe had changed the global landscape and a new set of plants and animals had appeared, with no particular connection with the previous flora and fauna.
4. *Inner-driven directional catastrophism* Louis Agassiz (1807–1873) espoused this position for most of his career. The underlying premise is that organisms have an immanent quality leading to progressive but discontinuous change. The progression of life was the unfolding of God's plan through catastrophes in the inorganic world and punctuations in the organic world, but with no causal connection between the two.

5. *Environmentally driven directional catastrophism* The mature William Buckland was a proponent of this position, that the Earth had suffered a series of catastrophes and that a new set of species was created after each mass extinction. Each new creation was an improvement on the previous one and the improvements placed the organisms in better harmony with the changed environment.
6. *Environmentally driven steady-state catastrophism* Baron Georges Cuvier was the leading advocate of this position. He could not accept a progression of organisms. He recognized four chief branches of animals, the members in each of which were fixed and designed to meet all environmental conditions. His catastrophes were essentially sudden environmental changes in the distribution of land and sea. The motor of biotic change is sudden environmental change.

The other two of Huggett's categories, the actualistic and inner-driven steady-state catastrophisms, were rarely expressed positions in the early nineteenth century but became more popular in the late twentieth century with the rise of neocatastrophism.

Since the 1960s, one of the key assumptions of catastrophism has been making a strong comeback in the form of neocatastrophism. This key assumption is that high magnitude, low frequency events are cumulatively more important in Earth history than low magnitude, high frequency events. Some reasons for this changed perspective are:

(1) Improved precision in geochronology has demonstrated unexpectedly rapid past changes.
(2) The exploration of mass extinctions in the past has intensified.
(3) Some geomorphological features, such as the channelled scablands of eastern Washington, are more amenable to explanation by low frequency, high magnitude events than by gradual, semi-continuous processes.
(4) Space exploration has generated a strong interest in galactic-scale events.
(5) Interest in global environmental change has provided evidence of rapid past changes, such as found in the polar ice caps and the oceanic deep sediments.
(6) The rise of non-linear dynamics and chaos theory is beginning to provide ways of synthesizing gradualism and catastrophism.

It seems self-evident that Earth history contains a combination of catastrophic and gradual events; accepting the occurrence of catastrophes is not to deny the effectiveness of gradual processes. The main reason for concerns about catastrophism expressed by Earth scientists since the mid-nineteenth century has been a fear of the reintroduction of religious beliefs into the canon of modern science because of the long-held views about the Noachian flood in western thinking. Largely for this reason, the geological establishment of the day refused to countenance the work of Harlen Bretz (1923) in his account of the origins of the channelled scablands of eastern Washington. He suggested that these SCABLANDS (massively and regionally gullied lands) could be explained best by the action of a single large flood over a period of only a few days. The fact that they have been shown subsequently to be caused by a succession of pulses associated with the draining of glacial Lake Missoula (Baker 1973) was complete vindication for Bretz. It is important to recognize that in no sense did the catastrophists violate the principle of uniformity of law (i.e. that natural laws are invariant in time and space). On the other hand, the non-actualists did violate the principle of simplicity, a principle which states that no unknown causes should be invoked if available processes are adequate. This guideline is known as the uniformity of process.

References

Baker, V.R. (1973) Palaeohydrology and sedimentology of Lake Missoula flooding in eastern Washington, *Geological Society of America Special Paper* 144.
Bretz, J.H. (1923) The channelled scablands of the Columbia Plateau, *Journal of Geology* 31, 617–649.
Chappell, J. (1978) On process-landform models from Papua New Guinea and elsewhere, in J.L. Davies and M.A.J. Williams (eds) *Landform Evolution in Australasia*, 348–361, Canberra: Australian National University Press.
Cuvier, G. (1817) *Essay on the Theory of the Earth*. With mineralogical notes an account of Cuvier's geological discoveries by Professor Jameson. Edinburgh: William Blackwood.
Gould, S.J. (1987) *Time's Arrow, Time's Cycle. Myth and Metaphor in the Discovery of Geological Time*, Cambridge: Harvard University Press.
Graf, W.L. (1979) Catastrophe theory as a model for changes in fluvial systems, in D.D. Rhodes and G.P. Williams (eds) *Adjustments of the Fluvial System*, 13–32, Dubuque: Kendall Hunt.
Huggett, R. (1990) *Catastrophism: Systems of Earth History*, London: Edward Arnold.
Thom, R. (1975) *Structural Stability and Morphogenesis*, New York: Benjamin.
Thornes, J.B. (1983) Evolutionary geomorphology, *Geography* 68, 225–235.

SEE ALSO: actualism; neocatastrophism; uniformitarianism

OLAV SLAYMAKER

CATENA

The concept of a catena is one of linkage – catena being the Latin for chain. So, by analogy individual elements are linked in some way and have something in common. In the case of a soil catena, the soils are linked by having the same parent material and age but are differentiated by their position on a slope which alone gives them different characteristics, especially in relation to drainage. Thus a formal definition is: 'A sequence of soils of about the same age, derived from similar parent material, and occurring under similar climatic conditions, but having different characteristics because of variations in relief and in drainage.'

A more generalized term which can be used is the toposequence, which refers to any sequence of soils which varies with topography. A more extensive, related concept is soil association, which is any repeated pattern in the landscape which may not necessarily be repeated with topography, as with a catena, but could be linked to geology, geomorphology and indicated by vegetation in any geomorphic element of the landscape.

The term soil catena was first proposed by Milne (1935a,b; 1936) for topographically linked soils in East Africa. On the sides of large valleys different soils were found on the upslope crest, the lower downslope and the footslope. Milne felt that the profiles of the soil types changed character downslope according to both drainage and the past history of the land surface. Here, there is an essentially uniform lithology throughout the slope and the differences derive from the shedding of moisture upslope and wetter conditions downslope as well as the movement of solutes in that water and the physical movement of eroded particles and their accumulation downslope.

Milne felt that the upper soils might also be older and more remnant and that the downslope soils might be younger, with a fresher accumulation of deposits, and also that the rocks might actually differ downslope. Ruhe (1960) further expanded on this by differentiating between the 'classic' catena on similar parent materials and a sequence which could be formed on two or more geological formations where the lithology actually varied. Here the downslope series is still linked by drainage and transport of material and shares a common physiographic history and geomorphic evolution despite being on different parent materials.

The catena concept is useful in soil mapping because it implies a regularly occurring relationship of the soils with topography, giving an expectation or prediction of what is liable to be present. An example of a general schema is in Table 6. With mechanical movement of particles, the upslope soils might be more coarse grained, having lost fines downslope, while the downslope soils could be more alluvial, with the accumulation of fines and/or with much coarser particles accumulating, according to the amount of mechanical action on the slope. Generally the midslope soils tend to be prone to erosion and are much thinner than the slope foot, where accumulation occurs.

The more complex and nutrient-rich montmorillonite clays may be found where nutrients accumulate at the base of the slope. In warmer areas, montmorillonite clays also survive on the soil crest and plateau but in wetter, cooler areas, the simpler kaolinite clays are found in these positions due to greater leaching. In hot, humid or semi-humid regions, the upslope leached soils are often red due to OXIDATION and at the slope foot where drainage is impeded the colour changes downslope through yellow to grey. The effect of topography can thus be expressed in the tropics with high rainfall where red soils with kaolonitic clays are found in the better drained upslope areas with montmorillonite and black, organic soils in the less well-drained depressions. On a larger scale, which includes tropical mountains, in areas of highland surrounded by lower arid land, the sequence with decreasing height is one of upland podzols and brown earths to chernozems to desert and semi-desert soils; for uplands surrounded by lower land with high rainfall the downslope sequence is one of podzols and brown earths to yellow, red and black soils of the humid lowlands.

The significance of a catenary sequence for agriculture and land-use planning is again predictive in that more drought resistant crops can be grown on the upper slopes and moisture-tolerant crops on the lower slope where the aerated zone is nearer the surface.

Hillslope hydrologists have also used the catena concept when devising predictive models for hillslope runoff generation. Here hydrological processes, such as infiltration, overland flow and throughflow, can be predicted in relation to soil profile properties which can be predicted to vary systematically downslope (McCaig 1985). Anderson (1985) also uses the concept of catena in predicting the mechanical, load-bearing properties of soils. Regression modelling has been attempted to quantify the relationship between slope position and soil properties with varying degrees of success (Furley 1971).

References

Anderson, M.G. (1985) Forecasting the trafficability of soils, in K.S. Richards, R.R. Arnett and S. Ellis (eds) *Geomorphology and Soils*, 396–416, London: George Allen and Unwin.
Furley, P.A. (1971) Relationships between slope form and soil properties developed over chalk parent materials, in D. Brunsden (ed.) *Slopes: Form and Process*, Institute of British Geographers, 141–163.
McCaig, M. (1985) Soil properties and subsurface hydrology, in K.S. Richards, R.R. Arnett and S. Ellis (eds) *Geomorphology and Soils*, 121–140, London: George Allen and Unwin.
Milne, G. (1935a) Some suggested units of classification and mapping, particularly for East African soils, *Soil Research* Berlin, 4, 183–198.
——(1935b) Complex units for the mapping of complex soil associations, *Transactions Third International Congress of Soil Science* 1, 345–1,347.

Table 6 The relationship of soils and topography

	Wetter, cooler climates	Wet climates	Drier climates
Upslope/ slope crest	Peat, podzolization	Podzolization	Leached soil
Midslope	Brown earth	Brown earth	Non-calcareous
Downslope	Peat, gley	Gley	Calcareous soil

Milne, G. (1936) A soil reconnaissance journey through parts of Tanganyika territory. *Memoirs of the Agricultural Research Station, Amani*. Reprinted 1947. *Journal of Ecology* 35, 192–265.
Ruhe, R.V. (1960) Elements of the soil landscape, *7th International Congress of Soil Science* 4, 165–170.

STEVE TRUDGILL

CAVE

The standard definition of a cave is 'an underground opening large enough for human entry' (Oxford English Dictionary and others). As such, natural caves occur in most consolidated rocks or compacted sediments and in most geomorphic settings, created by a variety of processes. This entry focuses upon KARST caves, which are the most important in terms of their magnitude, frequency, diversity of form, and role in general geomorphology. They are formed where dissolution is either the quantitatively predominant or essential trigger process of rock removal. Karst caves thus are restricted to the comparatively soluble rocks. In descending order of solubility, these are salt, gypsum and anhydrite, limestone and dolostone and (to a much lesser extent) quartzites, calcareous and siliceous sandstones. Limestone caves display the greatest range in size and form. Although 'enterable by humans' is the criterion, initial solution caves are much smaller (~1 cm diameter is a reasonable minimum) and the rules of genesis do not change significantly as they are enlarged to enterable size.

Caves created with little or no dissolution are PSEUDOKARSTic. CAVERNOUS WEATHERING voids can be considered transitional. Many pseudokarst caves are created by mechanical processes, including piping (see PIPE AND PIPING), THERMOKARST collapse, frost riving, wave action and CAVITATION along coasts. Such caves are rarely longer than a few tens of metres; most do not pass out of the range of daylight. Stream melting and sublimation in GLACIER ice and firn creates greater caves that are nearly identical in form and scale to the vadose shafts, canyons and simple phreatic passages found in many dissolutional caves. The lengthiest pseudokarst caves are lava tubes (see LAVA LANDFORM), formed by channelled discharge of still-molten lava within consolidating flows; single tubes, dendritic networks and anastomosing mazes are known, some extending for 10 km or more (Gillieson 1996).

Karst caves

Modern classification (Klimchouk *et al.* 2000) recognizes three principal genetic settings: (1) coastal caves in young carbonate rocks; (2) hypogene caves, formed by waters ascending out of artesian traps ('confined groundwaters') in any soluble rock; (3) unconfined meteoric water caves in soluble rocks, the most abundant and significant class. Figure 18 shows a basic range of plan patterns in these caves, relating them to type of recharge and the most effective (transmissive) porosity existing at the onset of dissolution. In most known hypogene and unconfined caves intergranular ('matrix') porosity is low, (<5 per cent), solvent water being transmitted via penetrable bedding planes, joints and faults which control the loci of the solutional passages. The relevant chemistry and kinetics of aqueous dissolution are discussed in Ford and Williams (1989: 42–126); Klimchouk *et al.* (2000: 124–223).

EOGENETIC COASTAL CAVES

'Eogenetic' describes very young limestone and dolostone accumulations where consolidation by compaction and interstitial cementation (i.e. diagenesis) is still limited, with the consequence that intergranular porosity offers principal or, at least, significant routes for solvent water flow. Such rocks are found in tropical/subtropical coastal settings today, e.g. Florida, Yucatan, Bahamas, many Pacific atolls, etc., and are chiefly Pleistocene in age. The matrix porosity yields cave patterns similar to those of cavernous weathering in non-karst strata.

'Syngenetic' caves (Jennings 1985) form in calcareous sand dunes when surficial sand becomes case hardened by cementation (i.e. earliest diagenesis), following which storm waves or surface streams breach the casing and wash out the non-cemented sand behind it, creating cavities sometimes many metres in length or height. This is one end of the spectrum of karst caves, where mechanical washout (piping) is quantitatively predominant but dissolution and cementation must precede it.

Much more widespread are caves formed where fresh and salt waters mix along the water table at the coast itself (flank margin caves) or at the halocline beneath the freshwater lens further inland (Klimchouk *et al.* 2000: 226–233). Flank margin caves display large entrance chambers dividing and tapering to blind endings a few metres or tens

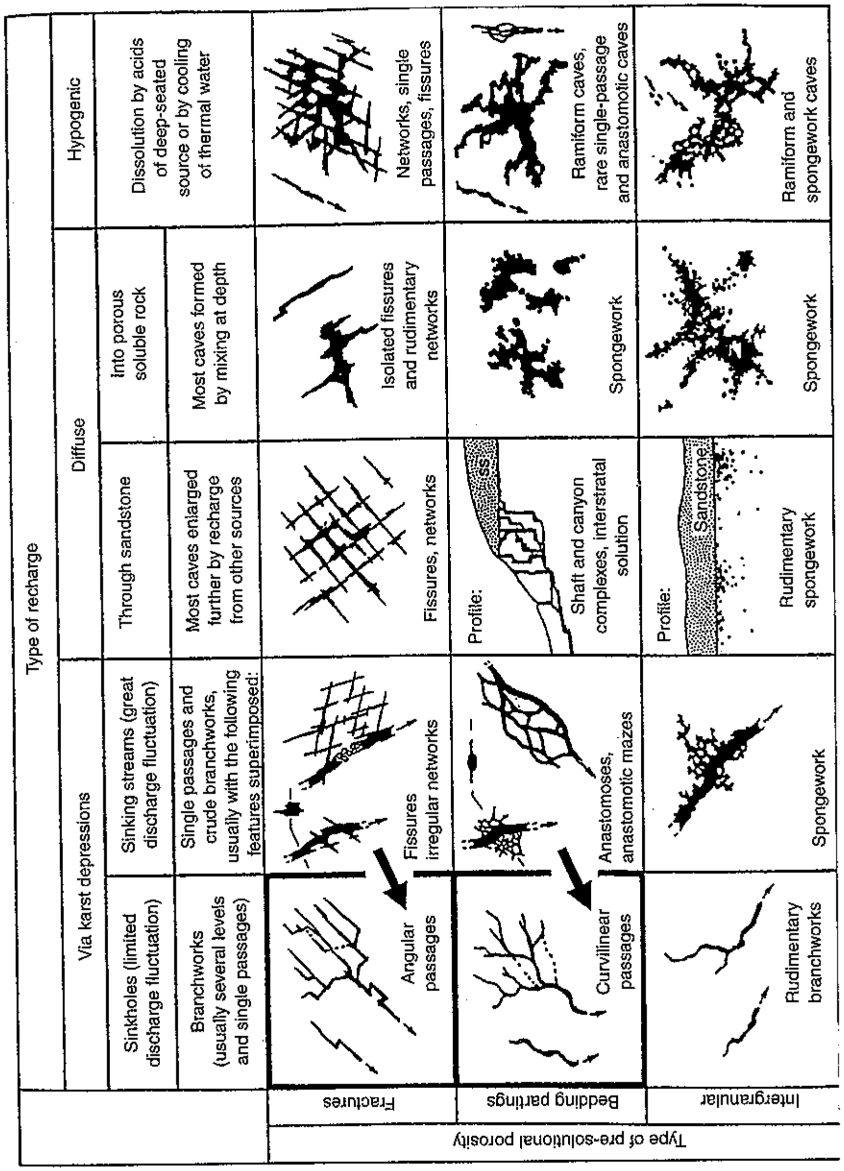

Figure 18 Basic plan patterns of caves shown in relation to types of pre-solutional porosity and conditions of recharge; slightly modified from Palmer (1991)

of metres inland. Halocline caves have more complex spongework patterns, in part due to the shifting of salt/fresh mixing zones as Quaternary sea levels moved up and down; aggregate passage lengths as great as 1 km are rare.

Cave systems many tens of km in length and extending 5 km or more inland are being explored along the Caribbean coast of the Yucatan Peninsula; although in young limestones, they are of unconfined meteoric origin, modified in form by the salt waters that now inundate them.

HYPOGENE CAVES

In hypogene caves the waters may have circulated deeply within karst strata alone (due to synclinal or graben-type structural traps), or be ascending into the karst rocks from underlying, non-karst aquifers (interstratal flow). They may be thermal (>4°C warmer than mean temperatures in the rock to be dissolved) or ambient.

A majority of hypogene caves are excavated under phreatic conditions, i.e. beneath any water table. The most simple form is a vertical/near-vertical chimney on a fracture, up which the water flows to discharge at a surface spring or into overlying, more porous, strata. Active instances include thermal springs in Mexico from shafts 40+ m in diameter and plumbed to −300 m. Deeper examples are known, but drained by uplift and erosion; some contain economic minerals precipitated on the walls, e.g Tyuya Muyun, Kazakhstan. More complex in form are arborescent chimney caves, branching upwards from basal reservoir chambers; Satorkopuszta Cave, Hungary, is a spectacular instance with later spheroidal rooms of condensation corrosion origin branching from an original phreatic shaft (Klimchouk *et al.* 2000: 292–303).

Fracture-guided network caves (Figure 18) are common. In western Ukraine local meteoric waters passing up through a ~14 m stratum of gypsum from an underlying sand aquifer created joint-guided mazes with intersections every 2–5 m; 212 km of contiguous passages are mapped in Opitimists' Cave. More complex are multistorey rectilinear mazes in the Black Hills, South Dakota, where thermal waters converged on Carboniferous palaeokarst (see PALAEOKARST AND RELICT KARST) preserved under clastic strata. The waters discharge through weaknesses in the clastics today, enlarging their routes and lowering the water table in Wind Cave 14 km distant at 40 cm kyr^{-1}. Jewel Cave, ~40 km distant, is fully drained; its 200 km of mapped passages are crusted with 10–20 cm of calcite spar deposited as the waters declined. Large-scale groundwater invasion and dissolution of limestones, gypsum and salt such as this, but deeply buried under later rocks, is associated with formation of solution breccias hosting oil and gas, lead/zinc and other mineral deposits, or creating breccia pipes that can stope upwards through 1,000+ m of overlying rocks. The greatest reported hypogene cavity is in Archean-Proterozoic marbles of the Rhodope Mountains, Bulgaria. It has an estimated volume of 237 million m^3 and a roof-to-floor depth believed to exceed 1,340 m (Klimchouk *et al.* 2000: 304–306).

A very distinctive type is the cave formed by sulphuric acid from H_2S. In most known instances the gas migrated from adjoining coal or oil basins where it was produced by bacterial reduction of gypsum. Reaching carbonate rocks it oxidized to H_2SO_4 at and just below the water table. Small H_2S caves tend to be linear outlets to springs. Large caves ramify about the gas inlets (Figure 18). Big chambers are created by lateral corrosion at the water table plus condensation corrosion above it that may convert limestone walls to gypsum to depths as great as one metre. Shift of inlet points and lowering of springs leads to multilevel development of spectacular systems such as Lechuguilla Cave, New Mexico (172 km, ~480 m in depth; Widmer 1998).

Unconfined caves

These are the principal type known to explorers and geomorphologists. The caves extend from surface water input points such as KARRENfields, DOLINES, river sinks or POLJES at the karst margins, to springs that are lower in elevation. Although it is rare for cavers to be able to follow the water all the way, dye tracing and other analysis invariably confirms that dissolutional conduits are continuous between sink and spring and regulate the flow. The most simple caves are single dissolutional pipes between sinks and springs. There are many instances in underground meander cut-offs or river short-cuts across narrow horsts or anticlines. However, the majority of caves have multiple inlets that, in plan view, link up to form crudely dendritic patterns. These are angular where joints are dominant controls and sinuous in bedding planes; many caves exhibit mixtures of the two. Joint mazes and bedding plane anastomoses (Figure 18) are usually subsidiary components in the dendritic plans, formed where there is rapid flooding at stream sinkpoints or underground blockages. The published plans of many cave systems appear more complicated than these simple combinations because the systems are multiphase; relict passages from higher levels cross those that are still active Modelling of plan pattern genesis is quite advanced, (see Ford and Williams 1989: 249–261; Klimchouk *et al.* 2000: 175–223).

Cave morphology in long section (i.e. length × depth) is closely related to geologic structure and groundwater zonation. In young mountain terrains if karst aquifers are thick and rapid uplift opens vertical fractures widely, caves are sequences of shafts down the steepest fractures, linked together by short, sinuous, stream canyons. Gravitational control of flow is predominant. A majority of the world's caves deeper than 1 km gain most of their depth in this upper vadose zone. Voronja Cave, Caucasus – currently the explored depth record holder at −1,710 m – is a fine example.

Where the karst formations are relatively thin and/or were little stressed during uplift, such conditions may never have existed. Instead, the uppermost zone is waterfilled initially but drains progressively as caves propagate through it and become enlarged – the 'drawdown vadose zone'. The caves display initial phreatic features such as elliptical cross sections in bedding planes, but with subsequent gravitational entrenchments beneath them. This combination can be found locally in young mountains also where passages are perched on shale bands or other obstructions. Entire cave systems, from sinkpoints to springs, can develop wholly within these two vadose settings, especially where karst strata are perched on insoluble rocks above regional base levels. There can be deep gravitational entrenchment into the insolubles; e.g. some 'contact caves' in Greenbrier County, West Virginia, have 90+ per cent of their volume in erodible shales beneath limestones that hosted the initial passages.

Most extensive cave systems have substantial water table or phreatic (sub-watertable) sections, however. Their length is usually greater than that of the vadose parts, and the geometry more complex and varied. There are four basic possibilities ('four state model'; Ford and Williams 1989: 261–274). If the density of penetrable, interconnected fissures is very low, geologic structure may compel the conduits to follow courses below the elevation of the springs or water table ('phreatic loops'). A State 1 system passes from the vadose zone to the spring in one loop. State 2 is a sequence of loops whose crests fix local elevations of the water table. Where fissure frequency and interconnection are greater, caves display mixtures of loops with gently graded passages at the water table (State 3). Very high frequency permits continuous development along the water table (State 4), similar to flank caves in young, porous limestones. There is probably phreatic looping to depths of 1,000+ m in some State 1 caves. Individual loops greater than 250 m deep are common in State 2. In regions subject to large magnitude, abrupt flooding such as alpine mountains, overflow ('epiphreatic') passages develop above the low stage conduits (Audra 1995).

Multi-level (multi-phase) caves

Most extensive caves have two or more 'levels' that developed to drain to successively lower springs – 'level' denoting the historical succession but not implying that the galleries must be near-horizontal; often, there is State 2 or 3 geometry. In vadose caves the lower levels may be simple entrenchments beneath the older passages. Where there is water table or phreatic development, the new springs are usually offset laterally tens to hundreds of metres or more and may have distributaries. The new springs steepen the hydraulic gradient in the downstream section of the old cave, which then adjusts to its new 'level' in a sequence of breakthrough undercaptures (French – *soutirages*) that move the hydraulic steepening progressively upstream like a river knickpoint. Portions of individual capture links are incorporated into the final dendritic pattern of the new level but others are left redundant, becoming drained relicts or silted backwaters.

Superimposition of successive levels, redundant links and invasion vadose caves from new sinkpoints, make maps of great systems such as Mammoth Cave, Kentucky (556 km – the longest mapped) more complex in appearance than almost any other geomorphic or hydrogeologic phenomena.

References

Audra, P. (1995) Karst alpins; Genèse des grands reseaux souterraines, *Karstologia Memoires*, 5.

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

Gillieson, D. (1996) *Caves: Processes, Development, Management*, Oxford: Blackwell.
Jennings, J.N. (1985) *Karst Geomorphology*, Oxford: Blackwell.
Klimchouk, A.V., Ford, D.C., Palmer, A.N. and Dreybrodt, W. (eds) (2000) *Speleogenesis; Evolution of Karst Aquifers*, Huntsville, AL: National Speleological Society of America.
Palmer, A.N. (1991) Origin and morphology of limestone caves, *Geological Society of America Bulletin*, 43.
Widmer, U. (ed.) (1998) *Lechuguilla, Jewel of the Underground*, Basel: Caving Publications International.

DEREK C. FORD

CAVERNOUS WEATHERING

Cavernous weathering is a process which causes the hollowing-out of rock outcrops and boulders on vertical and near-vertical faces. The hollows, or caverns, may take one of two forms. The first are known as 'tafoni' (singular: tafone), a term derived from the Sicilian word meaning 'windows'. TAFONI typically conform to a spherical or elliptical shape, have arched-shaped entrances, concave inner walls, overhanging visors and gently sloping debris-covered floors. Tafoni range in size from several centimetres to several metres in diameter and depth, may coalesce or interconnect and second order tafoni may develop on the back-walls of larger forms. The second form caverns may take is commonly called 'alveoli' (singular: alveole). Alveoli are formed by similar processes, termed HONEYCOMB WEATHERING, which involves the progressive development of closely spaced cavities in rock faces. These small-scale caverns are separated by narrow, intricate walls, creating a surface reminiscent of honeycomb. Alveoli are usually several centimetres in diameter and rarely are larger than one metre wide. The relationship between alveoli and tafoni has not been clearly defined and the distinction is therefore one of size and shape. Although cavernous and honeycomb weathering frequently occur together, their independent existence and differences in form have led some geomorphologists to contend that they are not genetically related, but rather derive from different modes of origin (Mustoe 1982). Cavernous weathering cannot be defined on the basis of geographical, lithological or climatological occurrence, as caverns may be found in many environments and on most rock types.

Caverns cannot be categorized on the basis of their occurrence on specific rock types or in specific climate regimes, as they occur under cold, temperate, hot, humid and arid environments, and are found on a variety of rock types. Cavernous weathering was previously considered to be a diagnostic feature of arid environments (Blackwelder 1929). Tafoni are most prolific in salt-rich environments, and have been documented most often in deserts (McGreevy 1985) and coastal zones (Mottershead and Pye 1994). Common factors in this disparate range of environments are high salt concentrations and frequent or occasional desiccating conditions. The occurrence of tafoni on sandstone surfaces (Young and Young 1992), and on granite surfaces (Dragovich 1969) has frequently been documented. Aside from these siliceous rock types, tafoni have also been recognized on tonalite, dolerite, lacustrine silts and conglomerates. A range of weathering processes may be responsible for the occurrence of tafoni, and there is clearly convergence of form. Considerable literature has accumulated on the nature of both tafoni and alveoli, but as more information has been presented, their possible origins, rather than being clarified, seem to have become more confused.

There are two types of tafoni: 'basal' and 'sidewall' tafoni. Basal tafoni are often found, as the name suggests, on outcrops and boulders at ground level. Sidewall tafoni are present on vertical and near-vertical outcrop surfaces where strong rock discontinuities are not present. Tafoni which develop along discontinuities, above ground level, may also be considered to be basal tafoni, and are characterized by a higher rate of back weathering than upwards progression.

Early studies of cavernous weathering suggested that caverns were created by the action of the wind, which was also responsible for the removal of weathering products. It is now widely accepted, however, that aeolian deflation is not responsible for the hollowing out of boulders or pitting of rock faces. Disintegration of cavern walls generally proceeds by flaking and granular disintegration, and numerous weathering processes have been invoked, including insolation weathering, WETTING AND DRYING WEATHERING, frost weathering (see FROST AND FROST WEATHERING), solution and chemical alteration of rock minerals, and SALT WEATHERING.

CHEMICAL WEATHERING processes are considered to be important in the development of tafoni in some circumstances. Tafoni in sandstone appear to be partly the result of the reaction of water and organic acids with iron and silica. Caverns in limestone are produced by a solution of calcium and magnesium carbonates; chemical weathering of dolerite and tonalite has been identified in caverns. Other chemical weathering processes which may have contributed to cavernous weathering include solution, HYDROLYSIS and hydration reactions, induced by microclimatological differences between caverns and exposed rock surfaces.

Of all the processes which may create caverns, salt weathering is the most commonly invoked. The importance of salt weathering is indicated by the presence of salt crystals on walls of coastal and desert tafoni and alveoli. Salts are evident in seepage, in granular debris and flakes being detached from cavern walls, in floor sediments and in crevices within caverns in many locations, testifying to the role of salt crystal growth in cavern development. Salt crystallization, hydration and thermal expansion may contribute to disaggregation, but given the wide geographical, climatological and lithological range of tafoni and alveoli, it is more likely that several weathering processes are involved in cavernous weathering.

A common feature of cavernously weathered surfaces is a case-hardened layer on exposed rock surfaces, penetrated by tafoni. Some researchers conclude that the presence of a hardened crust and weakened interior is a fundamental reason for tafoni occurrence (Conca and Rossman 1982). Conca and Rossman (1985) postulated that the presence of a case-hardening cement is of secondary importance compared to core softening, a result of differential weathering rates between the rock interior and exterior. Many other examples of tafoni in the absence of case hardening have been presented, however, so it appears that a hardened outer crust is not a prerequisite for tafoni formation.

A problem that has beset studies of tafoni weathering has been the tendency to relate it to one single formative process, whereas in many cases cavern development can only be satisfactorily explained by invoking the operation of a range of weathering mechanisms. There may be processes which are active in all cases, but it is likely that the relative importance of physical and chemical weathering processes will vary with different environmental conditions, which may operate under different catalytic conditions, and may act synergistically.

The relative significance of lithological controls is unclear in the context of tafoni origin. Caverns may be initiated along pre-existing joints or bedding planes, or may be distributed randomly across rock surfaces. This random (or pseudo-random) distribution may reflect points of mineralogical weakness on the rock surface, but this idea cannot be tested as the initial surface has been lost in the creation of the hollow.

Once initiated, the sheltering afforded by caverns may provide temperature ranges which are less extreme than on rock surfaces exposed to direct insolation, and higher relative humidities, a significant factor influencing the deposition, absorption and evaporation of moisture on rock surfaces. Microclimatological differences created in shadow zones of tafoni may accelerate the effects of weathering processes. Conversely, the exposed surfaces may shed moisture and solutes more rapidly, creating a negative, or self-regulating, feedback in the weathering system. Cavernous weathering may therefore represent the response of the weathering system to dynamical instabilities, where the positive feedback produced by material loss encourages accelerated weathering within caverns and the system responds by divergent evolution of the rock surface into hollows and exposed stable surfaces.

References

Blackwelder, E. (1929) Cavernous rock surfaces of the desert, *American Journal of Science* 17(101), 394–399.
Conca, J.L. and Rossman, G.R. (1982) Case hardening of sandstone, *Geology* 10, 520–523.
Conca, J.L. and Rossman, G.R. (1985) Core softening in cavernously weathered tonalite, *Journal of Geology* 93, 59–73.
Dragovich, D. (1969) The origin of cavernous surfaces (tafoni) in granitic rocks of southern South Australia, *Zeitschrift für Geomorphologie NF* 13, 163–181.
McGreevy, J.P. (1985) A preliminary Scanning Electron Microscope study of honeycomb weathering of sandstone in a coastal environment, *Earth Surface Processes and Landforms* 10, 509–518.
Mottershead, D.N. and Pye, K. (1994) Tafoni on coastal slopes, south Devon, U.K., *Earth Surface Processes and Landforms* 19, 543–563.
Mustoe, G.E. (1982) The origin of honeycomb weathering, *Geological Society of America Bulletin* 93, 108–115.
Young, R.W. and Young, A.R.M. (1992) *Sandstone Landforms*, Berlin: Springer-Verlag.

Further reading

Goudie, A.S. (1997) Weathering processes, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology* (2nd edn), 25–39, Chichester: Wiley.

Goudie, A.S. and Viles, H. (1997) *Salt Weathering Hazards*, Chichester: Wiley.
Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Clarendon Press.

SEE ALSO: case hardening; weathering

ALICE TURKINGTON

CAVITATION

A form of erosion that can occur in rapidly moving water. Local areas of low pressure may be created in the water and as Drewry (1986: 68) has explained:

> If the pressure falls as low as the vapour pressure of the water at bulk temperature, macroscopic bubbles of vapour (cavities) will form. The cavitation bubbles grow and are moved along in the fluid flow until they reach a region of slightly higher local pressure where they will suddenly collapse. If cavity collapse is adjacent to the channel wall localized but very high impact forces are produced against the rock. This action may give rise to mechanical failure of the channel.

The destructive action of cavitation is probably due to the shock waves created when the bubbles collapse. Cavitation is of great significance in the malfunction of hydraulic machinery (e.g. turbine blades, ships' propellors, etc.) but its geomorphological effects may also be substantial (Barnes 1956). Sufficiently high velocities for its operation can occur in such situations as waterfalls, rapids, bedrock channels, beneath glaciers and on TSUNAMI scoured surfaces (Aalto *et al.* 1999). Mean flow velocities necessary to initiate the process are generally higher than about $10\,\mathrm{m\,s^{-1}}$ for flow depths greater than about 4 m. Cavitation may contribute to the fluting and potholing of massive, unjointed rocks in bedrock channels (Whipple *et al.* 2000).

The term cavitation has a second and unrelated meaning in geomorphology, namely the formation of cavity at the bed or a sliding glacier. Their formation can enhance basal sliding (Lliboutry 1968).

References

Aalto, K.R., Aalto, R., Garrison-Laney, C.E. and Abramson, H.F. (1999) Tsumani(?) Sculpturing of the pebble beach wave-cut platform, Crescent City area, California, *Journal of Geology* 107, 607–622.
Barnes, H.L. (1956) Cavitation as a geological agent, *American Journal of Science* 254, 493–505.
Drewry, D. (1986) *Glacial Geologic Processes*, London: Arnold.
Lliboutry, L. (1968) General theory of subglacial cavitation and sliding of temperate glaciers, *Journal of Glaciology* 7, 21–58.
Whipple, K.X., Hancock, G.S. and Anderson, R.S. (2000) River incision into bedrock: mechanics and relative efficacy of plucking, abrasion and cavitation, *Bulletin of the Geological Society of America* 112, 480–503.

A.S. GOUDIE

CAY

Cays are the general terms for islands which develop on CORAL REEFS but can be divided for geomorphological purposes into true cays and *motu* (Nunn 1994). Both types are dominated by clastic materials scooped off the front of a coral reef, particularly off reef-talus slopes by large-amplitude waves, and dumped on reef surfaces. Such deposits have been observed to migrate across reef surfaces away from the oceans until they reach a point where they accumulate.

True cays are more transient, sometimes in existence for less than one year, compared to *motu*, some of which are 3,000–4,000 years old. In general, true cays are confined to narrow reef flats, commonly in either high-energy wave environments and/or in places affected annually by storm surges. *Motu* tend to be confined to broader reef surfaces, typically those outside the hurricane (tropical-cyclone) belt, where Holocene sea level exceeded its present level around 4,000 cal. yr BP.

True cays

Being ephemeral and transient reef islands, true cays are generally distinguished by being bare of vegetation and regularly overtopped by waves. They are also characterized by the absence of cemented deposits which renders them vulnerable to obliteration by large waves.

Although we know that true cays usually form as a result of storm surge (or tsunami) deposition of reef-front sediments, we do not clearly understand why this happens only in some instances while in others cays can be removed. It is likely that cay formation occurs when waves are coming from a direction where, in running up the submarine island slope, they can (and do) pick up a lot of material which is carried on to the reef surface. Cay removal may occur at higher velocities

but may also be when the wave picks up little material during run-up so that the main outcome of its passing across a reef surface is erosion rather than deposition. But there are other factors involved, particularly to do with reef morphology, sediment character and wave aspect which may lend a cyclical dimension to cay formation and removal. Studies of cays on Ontong Java Atoll in the western Pacific were an important step towards understanding this process (Bayliss-Smith 1988).

One characteristic of cays (and to a lesser extent of *motu*) is their mobility across reef flats. Historical data show that cays change shape regularly and even migrate across reef flats with erosion along the windward side commonly being compensated by progradation along the leeward side. A good example is that of Sand Island off St Croix in the Caribbean (Gerhard 1981). Such movements are the bane of cay-based tourist resorts.

Many cays endure for more than a few decades because they grow large enough to become vegetated and are in appropriate locations to develop beachrock. Such cays are better referred to as *motu*.

Motu

The main way in which *motu* can be distinguished from true cays is by the inclusion of shingle ridges within their fabric (Steers and Stoddart 1977). Such shingle ridges tend to be the residuals of rubble banks thrown up on the outer (ocean) sides of reefs by storm surges. As these ridges migrate landwards or lagoonwards, so the finer material is removed by wave wash so only the coarser fractions remain. Since they are so difficult to shift, particularly when located on the least vulnerable parts of a reef, these ridges often form the core of an atoll *motu*. The migration of the rubble bank thrown up on the Funafuti (Tuvalu) reef during Tropical Cyclone Bebe in 1972 was monitored by Baines and McLean (1976). Later work on the other atolls of Tuvalu demonstrated that such coarse shingle banks were integral parts of *motu*, particularly along windward reefs (McLean and Hosking 1991).

Motu also persist longer than true cays because they develop various forms of physical protection against wave erosion. These include emerged reef, BEACH ROCK, conglomerate platforms (*pakakota*) and phosphate rock, all of which greatly increase the resistance of *motu* against wave and/or precipitation attack, particularly during storms.

Those CORAL REEFS which were able to 'keep up' with Holocene sea-level rise grew to levels of 1–1.5 m above present levels in most parts of the tropics (except apparently the Caribbean) around 4,000 cal. yr BP (Nunn 2001). When the sea level fell by this amount in the later Holocene, the surface of these reefs was exposed and died. Subaerial erosion reduced them and wave erosion trimmed them, but many remained to act as foci for the accumulation of reef detritus. These fossil-reef cores underlie many *motu* today in the central Pacific, for example, and explain their persistence and suitability for human habitation.

Beach rock forms in a variety of ways beneath the surface of sandy beaches within the regularly inundated zone. For beach rock to form also usually requires a critical mass of sediment (equated with minimum *motu* size) so that ground water can flow through the beach sand.

Conglomerate platforms or breccia ramparts are cemented features of many *motu* thought to have formed at present sea level. Although there is clearly some unexplored genetic diversity amongst these features, most are believed to have originated as rubble banks which were subsequently cemented and planed down (McLean and Hosking 1991).

Phosphate rock also forms through the lithification of unconsolidated sediments, but on those *motu* where large numbers of seabirds roost (Stoddart and Scoffin 1983).

The future of cays

Many cays (including *motu*) have experienced apparently unprecedented morphological changes during the twentieth century, many of which can be attributed to the sea-level rise of ~15 cm. It is projected that twenty-first century sea-level rise may be 3–4 times as much, which has resulted in many gloomy prognoses for the future of cays (Roy and Connell 1989).

Should projections of sea-level rise prove correct, then it is likely that the numbers of cays worldwide will decrease hugely. First they may decrease because sea-level rise will, through the Bruun Effect, cause the erosion of sandy shorelines. It may become more common to see lines of beach rock exposed across reef flats marking places where cays once existed. Also, because of the rise of mean sea level and the likely inability of most oceanic reefs to respond immediately (despite some optimistic forecasts), it is probable that sediment of every grade presently lying on reef surface will become more mobile.

For many people occupying cays, particularly in independent countries like the Maldives (Indian Ocean) and Kiribati, Marshall Islands, Tokelau and Tuvalu (Pacific Ocean), it is unlikely that they will be able to continue living in such environments and will become 'environmental refugees'. Questions about national sovereignty and whether or not the Exclusive Economic Zones (EEZs) of these countries will be redrawn as a consequence are exercising the minds of many decision-makers.

References

Baines, G.B.K. and McLean, R.F. (1976) Sequential studies of hurricane deposit evolution at Funafuti atoll, *Marine Geology* 21, M1–M8.

Bayliss-Smith, T. (1988) The role of hurricanes in the development of reef islands, Ontong Java atoll, Solomon Islands, *Geographical Journal* 154, 377–391.

Gerhard, L.C. (1981) Origin and evolution of the Candlelight reef-sand cay system, St. Croix, *Atoll Research Bulletin* 242.

McLean, R.F. and Hosking, P.L. (1991) Geomorphology of reef islands and atoll motu in Tuvalu', *South Pacific Journal of Natural Science* 11, 167–189.

Nunn, P.D. (1994) *Oceanic Islands*, Oxford: Blackwell.

——(2001) Sea-level change in the Pacific, in J. Noye and M. Grzechnik (eds) *Sea-Level Changes and their Effects*, 1–23, Singapore: World Scientific Publishing.

Roy, P.S. and Connell, J. (1989) The Greenhouse Effect: where have all the islands gone? *Pacific Islands Monthly* 59, 16–21.

Steers, J.A. and Stoddart, D.R. (1977) The origin of fringing reefs, barrier reefs and atolls, in O.A. Jones and R. Endean (eds) *Biology and Geology of Coral Reefs*, Volume 4, 21–57, New York: Academic Press.

Stoddart, D.R. and Scoffin, T.P. (1983) Phosphate rock on coral reef islands, in A.S. Goudie and K. Pye (eds) *Chemical Sediments and Geomorphology*, 369–400, London: Academic Press.

SEE ALSO: coral reef

PATRICK D. NUNN

CENOTE

A cenote is a distinctive type of DOLINE or sinkhole, formed by the dissolution of limestone or other soluble rocks in subdued KARST plains (Plate 21). The type example occurs in the northern Yucatan Peninsula of Mexico, where the Mayan word dzonot, from which cenote is derived, means 'water cave'. Cenotes also occur in south-eastern Australia, in Africa, in Papua New Guinea, in Florida and in north-western Canada (Marker 1976).

The typical Yucatan cenote is a near-circular, water-filled shaft with vertical or overhanging walls extending up to 100 m downward from the ground surface (Pearse *et al.* 1936; Corbel 1959; Gerstenhauer 1968; Doering and Butler 1974). Some Yucatan cenotes resemble cylindrical shafts, but others are flooded bell-shaped chambers with bulbous bases, relatively small surface openings and thin roofs (Reddell 1977). Some have horizontal cave passages leading off from the walls, although these are often blocked by fallen rock (breakdown). The upper portions of cenote walls generally are pitted by dissolution, but the lower walls are blocky and overhanging, suggesting collapse (Whitaker 1998).

The development of cenotes is incompletely understood. Earlier hypotheses suggested that they developed through the local focusing of downward surface dissolution, but it now appears more likely that they have evolved through localized upward dissolution along fractures intersecting groundwater-filled caves or by stoping of cave ceilings, ultimately leading to surface collapse. Global sea-level oscillations have probably played a significant role too, since most cenotes are developed in Tertiary or younger reef limestones in low-lying coastal areas (Marker 1976). Sea-level lowering would have encouraged collapse by reducing the buoyant support of cenote rock walls and ceilings. Cenote-like flooded shafts, known as BLUE HOLES, occur in offshore reefs in the Caribbean and Australia (Mylroie *et al.* 1995). These may be drowned cenotes, although some of them have other origins (Ford and Williams 1989).

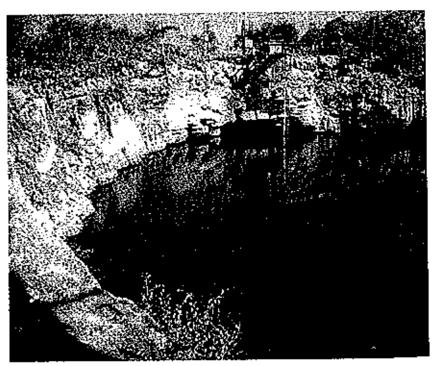

Plate 21 A steep-sided cenote formed in dolomitic limestones at Otjikoto near Tsumeb, in northern Namibia

The distribution of cenotes may be unrelated to other karst landforms, but they are often distributed in a linear pattern that may reflect underground fracture patterns or the paths of major cave passages. It has been suggested that the arcuate pattern of the Yucatan cenotes represents fracturing around the perimeter of the Chicxulub impact crater, which formed at the end of the Cretaceous period, some 65 million years ago (Hildebrand *et al.* 1995).

References

Corbel, J. (1959) Karst du Yucatan et de la Floride, *Bulletin de l'Association Géographique de France* 282/283, 2–14.

Doering, D.O. and Butler, J.H. (1974) Hydrogeologic constraints on Yucatan's development, *Science* 186(4,164), 591–595.

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

Gerstenhauer, A. (1968) Ein Karstmorphologischer Vergleich zwishen Florida und Yucatan, *Deutscher Geographisher*, 322–344, Bad Godesburg: Wissen Abhandlungen.

Hildebrand, A.R., Pilkington, M., Connors, M., Ortiz-Aleman, C. and Chavez, R.E. (1995) Size and structure of the Chicxulub crater revealed by horizontal gravity gradients and cenotes, *Nature* 376(6,539), 415–418.

Marker, M.E. (1976) Cenotes: a class of enclosed karst hollows, *Zeitschrift für Geomorpholgie N.F. Supplementband* 26, 104–123.

Mylroie, J.E., Carew, J.L. and Moore, A.I. (1995) Blue holes: definition and genesis, *Carbonates and Evaporites* 10, 225–233.

Pearse, A.S., Creaser, E.P. and Hall, F.G. (1936) *The Cenotes of Yucatan*, Carnegie Institute Publication 457. Washington, DC: Carnegie Institute.

Reddell, J.R. (1977) A preliminary survey of the caves of the Yucatan Peninsula, *Association for Mexican Cave Studies Bulletin* 6, 219–296.

Whitaker, F.F. (1998) The blue holes of the Bahamas: an overview and introduction to the Andros Project, *Cave and Karst Science* 25(2), 53–56.

MICHAEL J. DAY

CHANNEL, ALLUVIAL

Unconsolidated sediment deposited by rivers in subaerial settings is called ALLUVIUM and river channels formed of alluvium usually have a mobile boundary and are self-adjusting in response to changing conditions. An alluvial channel is commonly parabolic or trapezoid in cross section with adjacent roughly horizontal FLOODPLAINs that are inundated when the channel exceeds *bankfull capacity* (see BANKFULL DISCHARGE). Due largely to tributary contributions, channels generally increase in size and discharge downstream.

Alluvial channel morphology is the product of complex fluid mechanical processes. It was not until the late nineteenth and early twentieth centuries that river channels received widespread and detailed investigation when research was undertaken into partially self-adjusting canals built by the colonial British on the Indian subcontinent. The most important subsequent advances in understanding natural river-channel form and process originated from research in the USA by L.B. Leopold, M.G. Wolman, J.M. Miller, S.A. Schumm and their associates in the 1950s and 1960s.

Channel gradient and knickpoints

Channel gradients are the result of two broad controls. An *independent* gradient is imposed on the stream by antecedent VALLEY forms, products of geological and hydrological history. However, an adjustable and therefore *dependent* component develops from the interaction of channel discharge, width, depth, velocity, sediment size, sediment load, boundary roughness and path sinuosity. Mackin (1948) stated that a *graded* (or equilibrium) stream is one in which, over a period of years, the slope is delicately adjusted to provide, with available discharge and prevailing channel characteristics, just the velocity required for the transportation of the material supplied from upstream. Due to bedrock constraints (see BEDROCK CHANNEL), confined upland channels are generally not at equilibrium gradient, whereas in the middle or lower reaches the valley is wider and a channel can more readily adjust gradient by altering sinuosity and hence path length. Following this original emphasis on gradient, later work has shown that slope provides adjustments in concert with a variety of other morphological and hydraulic parameters (Leopold *et al.* 1964).

For three reasons the long profiles of natural rivers show a strong tendency for upwards concavity. First, downstream discharge increases as a cubic function whereas the resisting channel boundary increases only as a squared function, so if gradient did not decline the growing imbalance between impelling and resisting forces downstream would cause flow to accelerate rapidly. Second, because grain size commonly declines downstream, the equilibrium gradient required for sediment transport must also decline. Third, antecedent relief and potential energy conditions along a river from headwaters to mouth cause random-walk models to develop concavity as the *most probable* profile-shorter streams tend to have less concave profiles and streams with greater relief exhibit greater concavity.

Marked downstream increases in channel gradient are termed KNICKPOINTS and may reflect changes in bedrock erosional resistance, changes in sediment load from tributaries, tectonic activity, meander cutoffs, removal of LARGE WOODY DEBRIS (LWD), or base-level changes in the past. In confined valleys, knickpoints as concentrated zones of erosion can migrate considerable distances upstream. Unconfined channels, however, can adjust more readily by increasing sinuosity and thereby locally reducing knickpoint gradients.

Channel equilibrium and threshold conditions

Because alluvial channels are open systems with mobile and deformable boundaries, they have the ability to self-regulate to the imposed flow and sediment load. This reflects DYNAMIC EQUILIBRIUM, a condition first described for rivers in the nineteenth century by G.K. Gilbert. If one variable is altered, the others adjust in a way that minimizes the effect of the change and the system will return to something like its original condition (*homeostasis*).

While rivers in dynamic equilibrium generally resist change, Schumm (1973) has shown that an *extrinsic* THRESHOLD can be reached when a progressive change in an external variable triggers a sudden change in the system's response. At an imposed critical change of slope or sediment load, a meandering channel can change abruptly to a braided channel (Schumm and Kahn 1972). Similarly, a gradual and progressive increase in the flow velocity may suddenly achieve the threshold for sediment entrainment, after which the whole channel bed becomes mobile. Changes can be initiated intrinsically when, with no external change, one of the variables reaches a critical condition (an *intrinsic threshold*). A meander cutoff is an example where gradual, ongoing adjustments to equilibrium conditions prevail in a channel until an intrinsic threshold is reached.

Biota, soils and channel form

Prior to the middle Palaeozoic, subaerial erosion was dominantly physical and produced abundant coarse material forming mostly braided river channels. In the late Silurian and Devonian the evolution of terrestrial plant communities and associated soils greatly enhanced chemical weathering and the production of clays. The development of cohesive banks of muddy sediment and stabilizing root systems must have changed river channels dramatically. There is a growing appreciation of the importance and complexity of river–vegetation interactions. Particular attention has been given to the influence of within-channel vegetation and large woody debris (LWD) on flow resistance, and of bankline vegetation on bank strength and channel morphology.

The evolution of animals, including dinosaurs in the Mesozoic, has undoubtedly played a part in channel formation. Large mammals (e.g. American buffalo, African hippopotamus and domestic cattle) as well as smaller mammals (such as beavers) have been documented trampling sediments, creating paths down banks and damming channels, leading to channel avulsion and initiation.

Hydraulic geometry, regime theory and dominant discharge

Acceleration due to gravity acts to move water and sediment downslope while flow resistance opposes such motion. The interaction of these two forces ultimately determines the ability of flow to erode and transport sediment and to shape the boundary of an alluvial channel.

Flow velocity is usually fastest at or just below the surface near the centre of a straight channel and declines towards the bed and banks, the flow field deforming through river bends (Figure 19). A narrow deep channel usually exhibits a relatively gentle velocity gradient towards a fine-grained erodible bed, and directs relatively steep gradients to banks that are often cohesive, well vegetated and therefore erosion resistant. Wide shallow channels tend to exhibit erodible banks and coarse and/or abundant bedload that requires high shear stress for transport, braided rivers being a classic type.

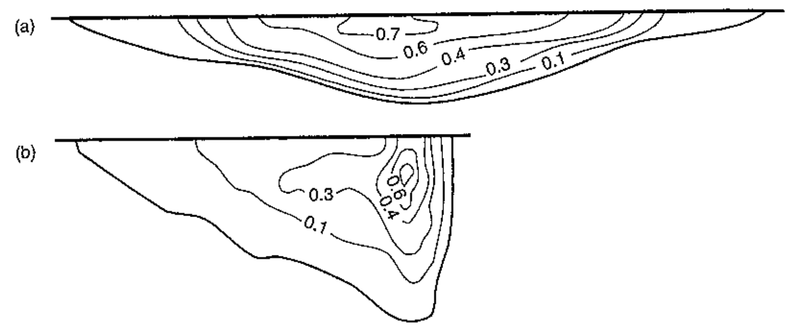

Figure 19 Velocity fields in cross sections of: (a) a wide shallow channel (note the steep velocity gradient to the bed); and (b) a narrow deep channel bend viewed downstream and curving to the left (note the steep velocity gradient against the outer cutbank)

Channel geometry is the cross-sectional form of a stream channel (width, depth, cross-sectional area) fashioned over a period of time in response to formative discharges and sediment characteristics. Because the above three geometric parameters and the additional four flow parameters (velocity, water-surface slope, flow resistance and sediment concentration) vary with discharge, the term HYDRAULIC GEOMETRY or *regime theory* is used to describe the relationships of all seven parameters to discharge as the independent variable (Figure 20). Consequently, stable alluvial rivers exhibiting consistent and predictable hydraulic geometries are said to be *in equilibrium* or *in regime*.

Discharge changes can be measured increasing at-a-station as the channel fills during a flood, or at bankfull in the downstream direction. There are significantly different relationships for *at-a-station* and *downstream* hydraulic geometry (Figure 20). Holding discharge constant, at-a-station hydraulic geometry is controlled mostly by variations in bank strength and available sediment load. Channels with low sediment loads and cohesive or well-vegetated banks tend to be relatively narrow and deep whereas those with abundant loads and weak banks tend to be wide and shallow. However, because bank strength has only a moderate range but river discharges vary by many orders of magnitude, hydraulic geometry is remarkably consistent across the full range of river discharges (Figure 20). Furthermore, because channel depth is greatly restricted by the limited strength of alluvial banks, rivers increase in width relative to their depth as their size and discharge increases – a prominent downstream tendency (Church 1992).

Hydraulic geometry shows that river channel dimensions are closely adjusted to water discharge. However, discharge varies from perhaps no flow in droughts through to catastrophic flood events, so which discharge(s) define a channel's characteristics? Leopold *et al.* (1964) showed that, in the USA, bankfull flows occur with the surprising regularity of about once every 1–2 years across a diverse range of rivers, something that would be an extraordinary coincidence if bankfull flows did not in themselves play a large part in determining channel dimensions. It has also been shown that with increasing at-a-station discharge, flow velocity tends to increase until near bankfull flow conditions and then stabilizes at higher discharges because of a marked increase in roughness near the bank crests and over the floodplains. In other words, most flows beyond bankfull are not notably more effective in altering the channel and transporting sediment than is bankfull flow. Furthermore, while in some cases exceptional floods may undertake significant work in the form of sediment transport and channel reconstruction, they are sufficiently rare that on an average annual basis, they usually achieve far less than do smaller but more frequent events of about bankfull stage (Figure 21) (Wolman and Miller 1960).

In some environments, particularly in confined alluvial settings, high-velocity events can cause considerable channel enlargement, followed by

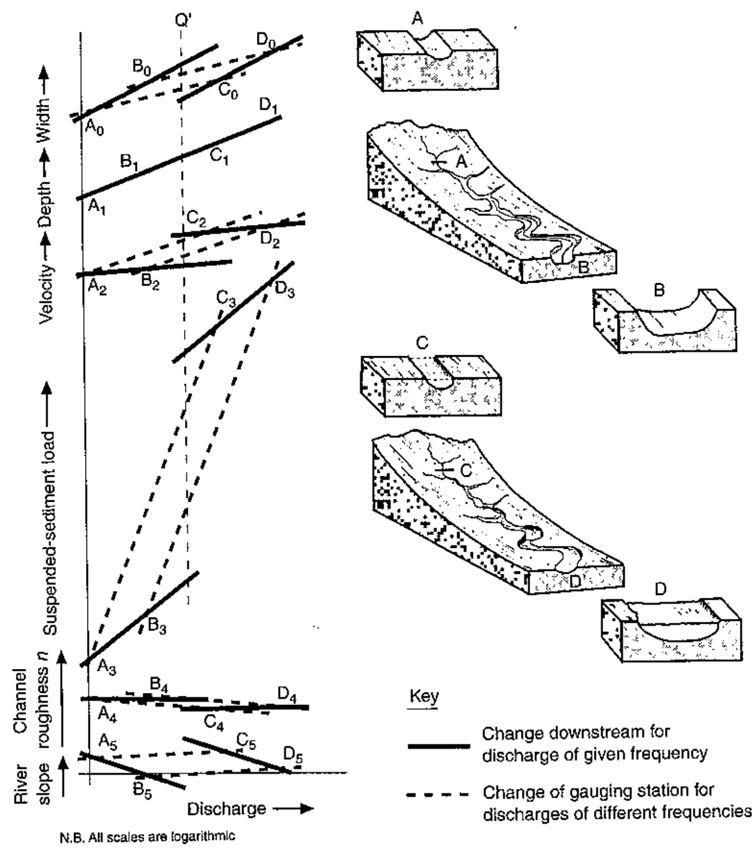

Figure 20 Hydraulic geometry relationships of river channels, comparing variations in width, depth, velocity, suspended load, toughness and slope to variations in discharge, both at-a-station and downstream (after Leopold *et al.* 1964)

a long period of 'recovery' from smaller flows. Thus, channel dimensions at a given time in such an environment may reflect considerable 'memory' of the last extreme event.

While empirical research into stochastic relationships has shown that, as flows vary, rivers construct highly predictable channel forms and sedimentary structures, a truly rational or deterministic explanation for such consistency has not been obtained. A lack of mathematical closure results from there being four flow variables (width, depth, velocity and slope) but only three determining equations (continuity, resistance and sediment transport). As a consequence, solutions have been sought by adopting *extremal hypotheses*, such as maximum sediment transport rate or minimum stream power. In a recent reassessment of some of these approaches, Huang and Nanson (2000) have demonstrated mathematically that straight reaches of alluvial rivers appear to operate at MAXIMUM FLOW EFFICIENCY (MFE) and illustrate the basic physical LEAST ACTION PRINCIPLE. However, although research into this principle is ongoing, such theoretical proposals remain contentious and are not uniformly accepted.

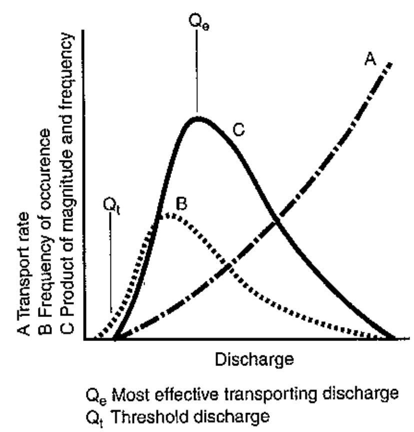

Figure 21 Dominant discharge. Curve A is the transport rate of suspended load rising with discharge. Curve B is the frequency of the full range of possible discharges. Curve C is the product of curves A and B and shows that the most effective discharges for transporting a river's load are moderate floods, generally occurring about once every 1 to 5 years (after Wolman and Miller 1960)

Channel patterns

River channels respond to imposed discharges and sediment loads by adjusting pattern or planform in conjunction with their hydraulic geometry. Because channel patterns are so easily recognized on air photos and maps they have become a primary basis for river classification from which it is possible to generalize less obvious channel characteristics such as lateral stability, sediment load, sediment size, bed/suspended load ratio and width/depth ratio.

Leopold *et al.* (1964) proposed the first widely adopted geomorphological classification with their concept of a continuum of river channel patterns from *straight* to MEANDERING to BRAIDED RIVERS, although a significant problem is that these are not mutually exclusive. For example, meanders sometimes also braid. Both meandering and braiding patterns appear to reflect a need to consume excess energy where the valley slope is greater than that required for an equilibrium channel slope (Bettess and White 1983; Schumm and Kahn 1972).

The term *anabranching* describes rivers that flow in multiple channels separated by stable, vegetated, alluvial islands that divide flows up to bankfull, regardless of their energy or sediment size, with *anastomosing* rivers simply a low-energy fine-grained type (see ANABRANCHING AND ANASTOMOSING RIVERS) (Nanson and Knighton 1996). Importantly, neither of these terms is now used as a synonym for braided rivers in which the flow is divided by unstable braid bars overtopped below bankfull.

Straight rivers consist of a single channel with a sinuosity of <1.1 (sinuosity is the ratio of channel length to valley length), a condition rarely persisting in an alluvial reach for distances of more than 7–10 channel widths. Consequently, straight reaches are classed as those without significant bends for more than this distance. Flume experiments suggest that straight channels form at very low gradients (Schumm and Kahn 1972). Where naturally sinuous channels in readily erodible material have been artificially straightened, alternate bars usually form rapidly and subsequent bank erosion leads to the development of a meandering pattern.

Braided rivers are relatively high-energy systems with large width–depth ratios and at low stage have multiple channels that divide and rejoin around alluvial bars. They tend to occur in settings with steep gradients, weakly cohesive banks, abundant coarse sediment (usually gravel and sand), and variable discharge (Leopold *et al.* 1964; Knighton 1998). Leopold *et al.* (1964) argued that braiding is an equilibrium adjustment to erodible banks and excessive load whereas Bettess and White (1983) see it as a pattern consuming energy in excessively steep valleys. Both explanations probably apply under different circumstances.

Meandering rivers consist of a single channel of moderate to low gradient with a sinuosity >1.3 and moderate width–depth ratios. Point bars commonly develop on the convex bank of a bend whereas the concave bank is typically erosional and adjacent to a pool. The locus of the lowest point in the channel (the thalweg) regularly oscillates laterally, switching channel sides at the riffle (crossover), a shallow zone in the long profile between each bend (pool). While there is no general agreement as to exactly how or why streams meander, they are self-similar over a wide range of scales. Width and wavelength in particular can be related to channel discharge (see Knighton 1998, Table 5.9). Using 'probability theory', Langbein and Leopold (1966) proposed that meanders reduce stream gradients to an equilibrium slope for the transport of an imposed sediment load (see also Bettess and White 1983), producing a longer path length with minimum variance and minimum total work. Meandering rivers tend to have cohesive and/or well-vegetated banks, mixed loads of sand (sometimes gravel) and mud, and commonly perennial flow. Channel lateral migration rates are most rapid where bends have a radius of curvature to channel width ratios of about 2 to 3 (Hickin and Nanson 1975).

Anabranching rivers are a system of multiple channels characterized by vegetated, stable alluvial islands which are either excised by avulsion from a previously continuous floodplain, or formed by the accretion of sediment in a previously wide channel. The islands divide flow up to bankfull (Nanson and Knighton 1996). Anabranching rivers include a wide range of subarctic, alpine, temperate, wet tropical and semi-arid settings and individual channels can be straight, meandering or braided. *Anastomosing* channels are low gradient, laterally stable, straight (most common) to highly sinuous variants, with low width–depth ratio and well-vegetated or highly cohesive banks. Anabranching rivers can confine flow and maintain an equilibrium bedload flux over low gradients, however, they can also distribute sediment over wide floodplains in disequilibrium, vertically accreting locations.

Church (1992) noted the problem of including rivers from mountains to basins within one classification scheme. He divided the full range of alluvial and non-alluvial channels into small, intermediate and large categories based not on channel dimensions but on the relationship between grain diameter (D) and depth (d). This approach offers opportunities for better classifying aquatic habitats but is less visually appealing.

Sediment transport and channel sedimentation

River channels transport their sediment load in essentially four ways; *bedload (traction load), saltating load, suspended load and dissolved load.* BEDLOAD is the coarsest fraction and moves short distances during relatively infrequent, high magnitude flows. It is commonly the smallest proportion of transported sediment (often <5 per cent of the total load), yet is of great geomorphic importance. It is largely bedload that controls channel configuration because its transport is a function of shear stress acting on the channel bed, and this is controlled by channel gradient (adjustable with sinuosity) and channel geometry. The capacity of alluvial rivers with unconstrained mobile boundaries to transport bedload is usually hydraulically defined, but few flows reach their capacity for transporting suspended load that is determined largely by the rate of supply. In other words, the character of a river is first determined by its imposed bedload, with suspended load and vegetation influencing bank cohesion and form. Because sediment character has a profound influence on river-channel morphology, Schumm (1960) developed a highly influential classification of rivers based on bedload, mixed load and suspended load systems, with width–depth ratios of >40, 10–40 and <10, respectively.

Alluvium results from fluvial sedimentation. This takes place both inbank and overbank as velocity wanes locally. The coarsest fractions are deposited first and as a result, sediment sizes are sorted vertically and laterally within the channel and floodplain. In laterally migrating meandering channels, upward fining successions within point bar and floodplain deposits result from flow velocities that decline from near the deepest part of channel (the thalweg) and adjacent point bar (depositing gravel or coarse sand), to the upper point bar and floodplain surface (depositing fine sand and mud). In braided rivers, coarse braid bars characterize the lowermost deposits while braid-channel and braid-bar migration or abandonment, overbank fines and channel fills characterize the uppermost deposits. Adjacent to laterally stable channels, or on floodplains away from the zone of active channels, floodplain strata broadly fine upward as each successive stratum makes the surface higher and less accessible to channel flows.

Secondary currents play a major role in producing the broad spatial variations of sediments in channel bends and bars, as well as numerous smaller flow structures. In gravel streams, prolonged flows near critical entrainment conditions can winnow fines and armour the surface, thereby lifting the threshold of bed motion during the next flood.

Conclusion

Alluvial channels represent continuum of forms that are classifiable on the basis of their cross-sectional shape, planform and associated processes. Whereas early research focused on stochastic relationships between channel form and process, there is a growing appreciation of rational explanations based on mechanics and accepted physical theory. Research into the operation and maintenance of alluvial channels remains a major area of pure and applied research within fluvial geomorphology.

References

Bettess, R. and White, W.R. (1983) Meandering and braiding of alluvial channels, *Proceedings of the Institution of Civil Engineers* 75, 525–538.

Church, M. (1992) Channel morphology and typology, in P. Calow and G.E. Petts (eds) *The River's Handbook: Hydrological and Ecological Principles*, 126–143, Washington, DC: Blackwell.

Hickin, E.J. and Nanson, G.C. (1975) The character of channel migration on the Beatton River, north-east British Columbia, Canada, *Geological Society of America Bulletin* 86, 487–494.

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.

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

Langbein, W.B. and Leopold, L.B. (1966) River meanders: theory of minimum variance, *United States Geological Survey Professional Paper*, 422H.

Leopold, L.B., Wolman, M.G. and Miller, J.P. (1964) *Fluvial Processes in Geomorphology*, San Francisco: Freeman.

Mackin, J.H. (1948) Concept of the graded river, *Geological Society of America Bulletin* 59, 463–512.

Nanson, G.C. and Knighton, A.D. (1996) Anabranching rivers: their cause, character and classification, *Earth Surface Processes and Landforms* 21, 217–239.

Schumm, S.A. (1960) The shape of alluvial channels in relation to sediment type, *United States Geological Survey Professional Paper* 352B, 17–30.

——(1973) Geomorphic thresholds and complex response of drainage systems, in M. Morisawa (ed.) *Fluvial Geomorphology*, 299–309, Binghamton, New York State University, Publications in Geomorphology.

Schumm, S.A. and Khan, H.R (1972) Experimental study of channel patterns, *Geological Society of American Bulletin* 83, 1,755–1,770.

Wolman, M.G. and Miller, J.P (1960) Magnitude and frequency of forces in geomorphic processes, *Journal of Geology* 68, 54–74.

SEE ALSO: armouring; bank erosion; channelization; confluence, channel and river junction; gravel-bed river; levee; long profile, river; models overflow channel; riparian geomorphology; river continuum; roughness; sediment load and yield; suspended load

GERALD C. NANSON AND MARTIN GIBLING

CHANNELIZATION

Channelization is the term used to describe the modification of river channels (usually alluvial channels, see CHANNEL, ALLUVIAL) by engineering. The aim is to provide flood control, improve land drainage, reduce erosion of channel banks and river beds, improve and maintain river navigation and to relocate channels in situations such as highway construction (see Brookes 1988 for a detailed text on channelized rivers). Some river channels have also been altered to float logs out from forests. River engineering can also create impounded rivers through the construction of DAMS (Petts 1984). Whilst the term channelization is extensively used, there are some equivalent terms used for the same group of engineering methods. These include 'kanalisation' in Germany, 'chenalisation' in France and 'canalization' in the UK.

River channelization has a long history and a large geographical coverage. Its origins can be traced to Mesopotamia and Egypt where there is evidence of river channelization for flood control and water supply as early as the sixth millennium BC. Indeed, most early civilizations constructed flood embankments. By 600 BC, reaches of the Huanghe (Yellow River) in China were embanked, and in Britain, the Romans constructed embankments to provide flood protection in the Fens and Somerset Levels.

Not surprisingly, the highest density of channelization occurs in developed countries associated with industrialization, urbanization and the intensification of agriculture. In the USA, 65 per cent of all channel alteration work is concentrated in Illinois, Indiana, North Dakota, Ohio and Kansas, with 51 per cent of all levee (embankment) work in California, Illinois and Florida (Brookes 1988: 10). In England and Wales, Brookes *et al.* (1983) estimated that for the period 1930–1980, 8,500 km of main rivers underwent major structural river engineering and a further 35,500 km were regularly maintained by dredging and weed cutting. And in Denmark, it is estimated (Brookes 1987) that 97.8 per cent of all streams were straightened by 1987. This is equivalent to a density of modified watercourses of 0.9 $km\,km^{-2}$ and compares with a density of channelized rivers in England and Wales of 0.06 $km\,km^{-2}$ (Brookes *et al.* 1983) and 0.003 $km\,km^{-2}$ for the USA (Leopold 1977). Thus, Denmark has a density of channelized river fifteen times greater than

England and Wales and 300 times greater than the USA, reflecting its intensity of land use.

Engineering techniques for flood control aim to prevent flood discharges overtopping the channel banks and spilling out onto the surrounding floodplain. Channels are designed and engineered to carry a design flood, which has a particular magnitude and frequency. If the 100-year event is selected as the design flood, the river channel will be engineered to contain and transmit a peak flow that will occur on average once every 100 years.

A range of 'structural' or 'hard' river engineering techniques are employed in channelization (Wharton 2000: 24–34) and many projects are comprehensive or composite in nature in that they employ more than one of the following engineering techniques.

Resectioning increases the cross-sectional area of the channel through widening and/or deepening. This allows flood flows that would have previously spread onto the floodplain to be contained and to flow through the channel at a lower and safer level. By combining with a process known as regrading (smoothing out the river bed by removing features such as depositional bars and pool-riffle sequences) the flow velocities are increased and flood levels are further reduced in the engineered reach.

Embankments, also known as levees, floodbanks and stopbanks, are structures built alongside a river to increase the bank height and prevent flooding onto the floodplain. They are normally constructed from material excavated from the channel or from a borrow pit in the floodplain, although imported materials are sometimes used. Detailed design specifications exist for embankments but a major consideration is the height, determined by the design flood.

Lining of channels in artificial materials is undertaken for both flood control and channel stability. Lined channels are common in urban areas and are normally rectangular in cross section with a straightened planform.

Realignment or straightening aims to improve the ease with which water flows through a river reach. The techniques range from removing deposited sediment by dredging (regrading), for example 'rock raking' carried out on gravel-bed rivers in New Zealand, to the removal of meander bends through cutoff programmes, for example the Middle Yangtze and Huanghe rivers in China and the Lower Mississippi river in the USA. River straightening also improves river navigation.

Diversion channels are relief channels constructed to divert flood flows away from an area requiring protection. The Jubilee River (completed in 2002) is a diversion or bypass channel providing flood relief for part of the River Thames catchment, UK. The newly engineered Jubilee River has a maximum capacity of $215\,m^3 s^{-1}$ and the main channel of the River Thames and existing right bank flood channels can carry up to $300\,m^3 s^{-1}$. It is predicted that the overall system capacity of $515\,m^3 s^{-1}$ will protect up to the 1 in 65-year return period flood. For environmental reasons the Jubilee River maintains a flow of $10\,m^3 s^{-1}$ at all times.

Culverts are structures that encase watercourses to provide flood protection. They may be masonry arches or large-diameter concrete or metal pipes. In many towns and cities, culverted streams flow beneath the streets, for example the rivers Fleet, Westbourne and Tyburn in London (UK).

Bank protection methods and river training works are engineering techniques for controlling river channel adjustments that could threaten settlements and agricultural land and have an impact on river navigation. Deposits from eroded riverbanks can also impede the river flow and increase the risk of flooding. Riverbanks have traditionally been protected by riprap (quarried stone), gabions (rock-filled wire baskets) and revetments (coverings of resistant materials such as concrete, steel or plastic sheeting). Although riprap is usually the preferred option, gabions do have an advantage in that the wire mesh allows the rocks to change position (caused by unstable ground or scouring of the riverbank) without failure. River training works are structures built to extend from the channel banks into the river and provide bank protection by deflecting erosive river flows away from vulnerable areas along the channel banks. The most common structures are groynes (also known as deflectors or dikes). Flows can be deflected onto channel deposits that pose problems for navigation or flood control to promote their removal through the natural process of scouring. Groynes have been used in this way on the Mississippi River to maintain a navigation channel. River training works can also be used to promote sediment trapping and deposition in areas that have suffered erosion. For example, a series of permeable groynes will allow water to pass through the structures but induce deposition of fine suspended sediment between the groynes, whereas impermeable groynes will deflect river flows and promote the trapping of larger bed material.

Dredging, weed cutting, clearing and snagging (collectively known as channel maintenance activities) are routinely undertaken on many rivers to improve the efficiency of water flow through the channel and reduce the flood risk. The removal of 'obstructions' to flow, reduces channel roughness, increases river flow velocity and lowers the flood height for a given discharge. Dredging may simply involve breaking up and loosening material for the river to transport downstream. In contrast, sediment may be removed by mechanical diggers, pumped onto the floodplain or be discharged into river barges before being dumped at selected locations. Weed cutting is practised in many streams, especially nutrient-rich chalk streams, to control the annual growth of submerged and emergent aquatic plants. In addition to physically reducing the capacity of the channel, plants also increase flood risk by increasing the resistance to flow and reducing water velocities. This further promotes the accumulation of sediment within and around the plants. Aquatic vegetation is traditionally controlled by mechanical cutting, but herbicides and grazing fish (such as carp) have also been used. Clearing and snagging refers to the removal of fallen trees and debris dams from the river and the harvesting of timber from the channel banks and floodplains, respectively.

A number of concerns surround river channelization. First, channelization, is unable to provide complete protection against flooding and its associated channel form adjustments. It is simply not possible or economically viable to control the very rare, high-magnitude flood events. To achieve this, all rivers would need to be channelized and all flood defences would need to be designed and constructed to convey a correctly estimated maximum possible flood. Second, there is evidence from developed countries with a long history of channelization that the financial costs of floods are continuing to rise despite ever-increasing expenditure on structural flood defences. This has been attributed, at least in part, to the false sense of security created by flood defences that encourages further floodplain development. And third, river channelization has resulted in changes to the river, many of which were not anticipated at the design stage. These changes can have a damaging impact on the river environment and also necessitate costly maintenance activities to keep the structures operating at their design specifications. Brookes (1988: 81–185) provides a comprehensive review of the main impacts of river channelization. Included in this third set of concerns are fears that river engineering may have worsened flooding on some rivers. Whilst channelization can reduce flood risk in the engineered reach, the reverse may be true downstream.

Brookes (1988) describes the primary impact of channelization as the physical alteration to the river (i.e. its width, depth, slope and planform) by the engineering procedure. These changes then result in secondary effects that encompass changes to the river channel morphology, hydrology, water quality and ecology. Importantly, these impacts are transmitted beyond the channelized river section to downstream and upstream reaches and even along tributary streams. Post-engineering adjustments demonstrate the need for long-term and often costly maintenance operations and also have implications for structures built adjacent to, or across, river channels. For example, bridges may have to be reinforced or even replaced if bank erosion causes the river to migrate and enlarge.

The reporting of channelization impacts and the appraisal of channelization schemes will lead to improved understanding of the various changes that river engineering may cause. Greater recognition of the undesirable consequences of channelization has led to calls for a 'reverence for rivers' (Leopold 1977) with attempts to design with nature (after McHarg 1969) and develop 'geomorphic engineering' (Coates 1976). This has translated into a variety of revised construction and maintenance procedures (see Brookes 1988: 189–209) and the development of more environmentally sensitive flood alleviation schemes, such as the flexible two-stage channel constructed on the River Roding, UK (Raven 1986). In this design, the additional capacity is created by excavating outside the original channel thus leaving it to transport the normal range of flows and remain as natural as possible. Growing concern over the impacts of channelization has also prompted efforts to enhance, rehabilitate and restore river systems (see RIVER RESTORATION).

References

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

Brookes, A (1988) *Channelized Rivers: Perspectives for Environmental Management*, Chichester: Wiley.
Brookes, A., Gregory, K.J. and Dawson, F.H. (1983) An assessment of river channelization in England and Wales, *Science of the Total Environment* 27, 97–112.
Coates, D.R. (ed.) (1976) *Geomorphology and Engineering*, London: George Allen and Unwin.
Leopold, L.B. (1977) A reverence for rivers, *Geology* 5, 429–430.
McHarg, I.L. (1969) *Design with Nature*, New York: Doubleday.
Petts, G.E. (1984) *Impounded Rivers: Perspectives for Ecological Management*, Chichester: Wiley.
Raven, P.J. (1986) Changes of in-channel vegetation following two-stage channel construction on a small rural clay river, *Journal of Applied Ecology* 23, 333–345.
Wharton, G. (2000) *Managing River Environments*, Cambridge: Cambridge University Press.

SEE ALSO: anthropogeomorphology; bankfull discharge

GERALDENE WHARTON

CHAOS THEORY

Chaos theory has been claimed by some enthusiasts as being one of the great ideas of twentieth century science which, as with relativity and quantum mechanics, has the power to transform our view of the world. As the popular book on chaos by James Gleick (1987) illustrates, chaos theory developed in a series of often unrelated spheres of science as developments in computing power permitted the increasingly sophisticated study of non-linear systems. Non-linear systems are those in which a change in one variable produces a non-linear response in another, and thus have to be represented by non-linear equations. Chaos is a property sometimes exhibited by such non-linear systems where even under simple conditions the system can tend to complex, pseudo-random behaviour. A classic paper by Edward Lorenz in 1963 illustrates the potential for chaos in relatively simple systems. Lorenz developed a simple climatic model of the atmosphere heated from below to produce convection, involving three non-linear equations. The three equations describe the change in x, y and z over time respectively, where x describes the intensity of convective motion, y the horizontal temperature variation and z the vertical temperature variation. Despite its simplicity the modelled system exhibited chaotic behaviour, indicating the unpredictable behaviour of this sort of climatic system.

Chaotic behaviour can be identified in systems through using phase diagrams, which plot the state of the system over time in terms of the system variables. In the case of the Lorenz model above, for example, the phase diagram would plot each point in time of the evolution of the system on x, y and z co-ordinates. A stable system would have a phase diagram which converged on a point, an oscillating or periodic system would have one which resembled a ring. Such shapes on a phase diagram are called attractors. Phase diagrams for chaotic systems are characterized by what are called 'strange attractors' – bifurcating, complex patterns illustrating the many different possible states of the system as it evolves over time. Lorenz's model, for example, has a strange attractor which looks like an owl mask. Strange attractors are fractals (see FRACTAL).

According to Malanson *et al.* (1990) chaos theory has three central tenets. First, that many simple deterministic systems are rarely predictable. Second, that some systems show great sensitivity to initial conditions. Tweaking an input to one equation of a system very slightly at the beginning can thus produce highly divergent outcomes. Third, that the conjunction of the first two tenets produces a seeming randomness which may be quite ordered (as illustrated by the presence of strange attractors in their phase diagrams).

Geomorphologists have been keen to investigate the utility of chaos theory ideas and methods for the study of geomorphic systems, many of which can be shown to be non-linear in nature. For example, in a series of papers Jonathan Phillips has investigated the presence of chaos in surface runoff, hillslope evolution, coastal wetlands and soil systems as reviewed in his book on Earth surface systems (Phillips 1999). Mass movement systems often behave chaotically (Qin *et al.* 2002). Increasingly, geomorphologists suspect that chaotic and self-organized behaviour (see SELF-ORGANIZED CRITICALITY) may be common within Earth surface systems, and that stable states may be relatively uncommon. However, chaotic behaviour may also be scale-dependent, and at other scales ordered behaviour may emerge. For example, at the microscale turbulence (a classic manifestation of chaotic behaviour) characterizes many aeolian-sediment interactions within dunefields, whereas at the larger scale ordered dune systems result. As Phillips (1999: 71) puts it 'Order is an emergent property of the unstable, chaotic system'.

Although chaos theory has undoubtedly stimulated much interesting and useful research and discussion in geomorphology, its application to geomorphic systems is not problem-free. Three key issues are discussed by Baas (2002). First, the presence of random noise within many geomorphic systems can often mask chaotic behaviour and make it almost impossible to analyse what is going on. Second, analysing chaos requires good datasets, which are not necessarily forthcoming in many areas of geomorphology, although the advent of good quality DIGITAL ELEVATION MODELS (DEMs) at a range of scales has started to help enormously in this regard. Finally, there are a range of different interpretations of chaos theory in the scientific literature, and many different methods available to analyse chaotic systems – all of which can be rather confusing to geomorphologists wishing to understand and utilize chaos theory.

References

Baas, A.C.W. (2002) Chaos, fractals and self-organization in coastal geomorphology: simulating dune landscapes in vegetated environments, *Geomorphology* 48, 309–328.
Gleick, J. (1987) *Chaos*, Harmondsworth: Penguin.
Lorenz, E.N. (1963) Deterministic non-periodic flows, *Journal of Atmospheric Sciences* 20, 130–141.
Malanson, G.P., Butler, D.R. and Walsh, S.J. (1990) Chaos theory in physical geography, *Physical Geography* 11, 293–304.
Phillips, J.D. (1999) *Earth Surface Systems: Complexity, Order and Scale*, Oxford: Blackwell.
Qin, S., Jico, J.J. and Wang, S. (2002) A nonlinear dynamical model of landslide evolution, *Geomorphology*, 43, 77–86.

Further reading

Malanson, G.P., Butler, D.R. and Geograkakos, K.P. (1992) Nonequilibrium geomorphic processes and deterministic chaos, *Geomorphology* 5, 311–322.
Sivakumar, B. (2000) Chaos theory in hydrology: important issues and interpretations, *Journal of Hydrology* 227, 1–20.
Turcotte, D.L. (1992) *Fractals and Chaos in Geology and Geophysics*, Cambridge: Cambridge University Press.

HEATHER A. VILES

CHELATION AND CHELUVIATION

Organic compounds, derived through the partial decomposition of organic matter, are important agents in weathering. Some act because they are acid and simply etch into minerals but for others, ions from the mineral actually become incorporated into the chemical structure of the organic compound and it is these compounds which are called chelates. The word is derived from the Latin *chela* and Greek *khele* which means a claw – and it can be readily imagined how the claw of, say, a crab can hold an object in the tips of its pincers and this is analogous to the way in which the chemical compound holds an atom derived from a mineral. The definition of a chelate can now be appreciated: 'a compound containing a ligand (typically organic) bonded to a central metal atom at two or more points'; where a ligand is: 'an ion or molecule attached to a metal atom by co-ordinate bonding'.

In the context of weathering, the metal ions of interest are commonly iron but can be zinc, copper, manganese, calcium or magnesium. Chelation weathering is then the process of the incorporation of these metal atoms into an organic compound derived from the decay of organic matter. The significance of this process is that many minerals are subject to weathering by chelates to a much greater degree than they are in water, even acidified water (Huang and Keller 1972; Huang and Kiang 1972).

Cheluviation is a compound word derived from chelation and eluviation (see ELUVIUM AND ELUVIATION). Since eluviation is the down-washing of material through the soil in mobile soil water, cheluviation involves the down-washing of chelates, with their associated metal cations, from the upper horizons of the soil to the lower horizons. It is in this way that iron can be moved from the upper horizons of a podzolic soil, rendering it a pale colour with an absence of reddish oxidized iron, ferric iron or Iron III. The process involves simultaneous chelation and REDUCTION of the iron to the mobile ferrous (Iron II) form. The iron then may accumulate lower down in the soil as a reddish or, because of the presence of organic matter, blackish layer. Here the reddish oxidized Iron III forms as a result of OXIDATION through a rise in pH which occurs in the lower parts of the soil profile which are less acid than the upper parts which are acidified by organic acids. The redeposition of the iron can be in the form of a hard iron pan, termed a BFe horizon, or a more diffuse reddish horizon. The latter is termed a Bs horizon as it contains sesquioxides which are defined as compounds such as Fe_2O_3 which have a ratio of metal to oxygen of $1{:}1\frac{1}{2}$.

References

Huang, W.H. and Keller, W.D. (1972) Organic acids as agents of chemical weathering of silicate minerals, *Nature (Physical Science)* 239, 149–151.

Huang, W.H. and Kiang, W.C. (1972) Laboratory dissolution of plagioclase feldspars in water and organic acids at room temperature, *American Mineralogist* 57, 1,849–1,859.

STEVE TRUDGILL

CHEMICAL DENUDATION

Central to any understanding of landform change through time is an understanding of DENUDATION rates (the volume of rock removed from a given area in a specified period of time). Knowledge of denudation rates is relevant also to geochemical and sediment mass balance studies, with important implications for global carbon budgets and global climate change. Denudation results from the removal of solid particles (mechanical denudation; Meybeck 1987) and dissolved material (chemical denudation). Overall, chemical denudation has received less attention than mechanical denudation and estimates of its local, regional and global significance often are subject to much uncertainty. The processes of CHEMICAL WEATHERING through which atmospheric, hydrologic and biologic agents act upon and alter mineral constituents of rocks by chemical reactions, thereby releasing dissolved material to be removed, are considered elsewhere. Here, an overview of the methods for studying chemical denudation and the variability of rates from different environments is provided. The role of relief, lithology and climate as controlling factors are discussed.

Methods for studying chemical denudation

SOLUTE YIELDS

Most frequently chemical denudation is calculated from the solute loads of rivers draining large catchments. An estimate of chemical denudation can be achieved simply by multiplying the mean solute concentration from samples of river water by mean discharge. More accurate estimates, however, take into account solute concentration relationships with discharge, particularly through floods using solute rating curves (see SOLUTE LOAD AND RATING CURVE) constructed from equations which best fit the relationship between solute concentrations and discharge. For greater accuracy these rating curves can be used for the rising and falling limbs of flood hydrographs. Solute transport rates can then be calculated by relating the rating curve to either continuous stream-flow data or flow-duration curves based on hourly, daily or even monthly data.

Given the complexity of measuring separately each dissolved constituent in stream water, electrical conductivity (specific conductance) of the water, which is more easily measured, is often used to provide an estimate of solute concentration. Although there is a strong correlation between the concentration of ionic species in solution and electrical conductivity, the exact relationship varies depending on concentrations present of particular dissolved constituents. Moreover, SiO_2, which may be a significant component of many tropical lowland rivers, is not recorded by this technique.

The most significant problem in estimating the contribution of solute transport to denudation is the separation of denudational and non-denudational contributions. Chemical weathering is not the only process affecting solute yields (Figure 22). Dissolved constituents introduced into a catchment from atmospheric wet and dry deposition must be accounted for. These atmospheric deposition fluxes can be quantified by direct measurement, though results often are highly variable with complex spatial patterns in regions with different vegetation types (Drever and Clow 1995). Global estimates of non-denudational atmospheric inputs (details in Summerfield 1991) from precipitation (oceanic salts) and atmospheric CO_2 (incorporated during weathering reactions), average approximately 40 per cent of catchment solute yields. The fraction is highest for the ions of Na, Cl and HCO_3 (50 to 70 per cent), although it is important to caution that these values vary greatly.

Further complications, depending on the timescale of study, relate to changes in the exchange pool of cations and anions in the soil and biomass (Figure 22). In the short term, changes in the soil occur as a result of precipitation events, evapotranspiration and the growth cycles of plants. As plants grow, they extract inorganic nutrients from the soil solution and incorporate them into plant tissue. When plants die and decompose, the process is reversed and the elements are returned to the soil. If an ecosystem is in steady state, new growth is exactly balanced by the death and decay of old vegetation and the biomass is neither a net source or sink. However, in forested catchments the biomass is rarely in steady state. For example, data of Likens *et al.* (1977) in Hubbard Brook indicated that the uptake of Ca by biomass was 45 per cent of the amount released by weathering. For K the value was 86 per cent.

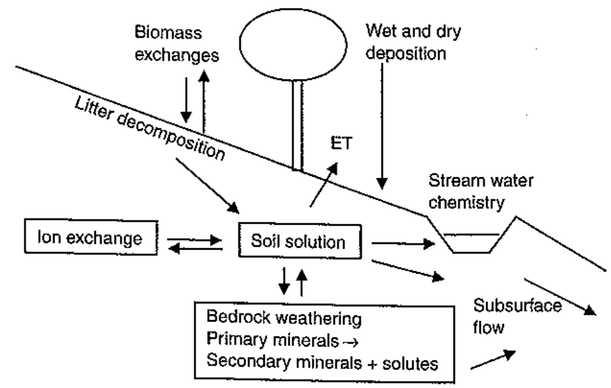

Figure 22 Schematic representation of key factors influencing solute fluxes in a catchment (adapted from Drever and Clow 1995)

The solute loads of rivers increasingly are being impacted by human inputs, especially where industrial and agricultural activities are concentrated. Elevated acid inputs increase the rate of chemical weathering in watersheds underlain by reactive rock types and cause acidification of water and soil in catchments underlain by non-reactive rocks types. Significant changes in the soil and biomass exchange pools can result, enhancing rates of solute input into stream waters. Such anthropogenic influence further complicates interpretation of solute concentrations in terms of 'natural' chemical denudation rates.

While corrections can be made to solute yields to account for atmospheric, biogenic and anthropogenic processes, the actual volume change (denudation) in the catchment cannot be determined unless the alterations in bulk density that accompany the weathering reactions releasing the solutes are known. While some chemical weathering dissolves bedrock minerals completely with all the products as dissolved species (common with limestone and quartzite) (congruent reactions), many weathering reactions produce both dissolved species and new solids (often clay minerals) with a similar volume but decreased bulk density. Moreover, even the congruent reactions and associated chemical losses may result in a substantial decrease in density as silicate rocks are altered to SAPROLITE (a weathered residuum retaining the structure and layering of the bedrock from which it forms) (density of bedrock 2.5–2.7; saprolite 1.3–1.7; soil 0.8–1.3). Thus there may be no discernible effect on the configuration of the landscape and direct conversion of dissolved loss into surface lowering is unrealistic.

SOIL PROFILE DEPLETION AND MASS BALANCE MODELLING

Soil mineralogy represents the residual product of chemical reactions which integrate the weathering rate over the entire period of soil development. Thus an alternative approach to determine long-term rates of chemical weathering and denudation is to quantify element and mineral losses in a soil profile relative to the initial or parent material. The most common approach is to define the mass ratio (enrichment) of a conservative component whose absolute mass does not change during weathering. As relatively soluble minerals in soils dissolve away, the more immobile elements in soils become increasingly enriched relative to their concentrations in the unweathered parent material. Measurement of enrichment of immobile elements, such as Zr, Ti, of rare Earth elements such as Nb, can reveal the degree of soil weathering and thus can be used to quantify the total dissolution loss from a soil (see examples in White 1995). However, there is considerable disagreement in the literature about the relative mobility of elements in different weathering regimes, and minor elements, such as Zr and Ti, are often concentrated in the small size, heavy fraction which may be subjected to significant fractionation during sediment transport and deposition. Assuming these are not major issues, the average weathering rate can be estimated by dividing the dissolution loss by the soil age. However, because non-eroding soils of known age are rare, this mass balance approach cannot be used in many environments. Riebe *et al.* (2001) show how the soil mass balance approach can be extended to measure long-term weathering rates in eroding landscapes. Physical erosion rates can be inferred from cosmogenic nuclides (see COSMOGENIC DATING), with dissolution losses inferred from rock-to-soil enrichment of insoluble elements.

Regional and global patterns

The distribution of studies of solute loads and chemical denudation is uneven. Most recent work on the rates and significance of chemical weathering in small watersheds has been driven by interest in the effects of acid deposition. Thus numerous studies have been conducted in North America and Europe, yet few data collected at the watershed-scale exist from other parts of the world. Thus estimates at continent-wide or global-scales must be extrapolated. This is usually based on empirical relationships observed between solute transport rates and the factors thought to control these rates, notably rock-type, climate and relief (see further discussion below). Moreover, records of dissolved loads tend to be short and results variable through time. Thus there are also problems extrapolating such data to the longer time periods over which ecosystems, soils, landscapes and climates evolve.

Some of the earliest regional estimates of chemical denudation were attempted by Dole and Stabler (1909) for the United States. Data compiled by Summerfield (1991) are used here to provide a range of estimates of solute load transport and equivalent rates of chemical denudation (see regional summary in Table 7). These estimates are subject to all the errors described above.

The data yield a global average for chemical denudation of $3{,}700\,\mathrm{Mt\,a^{-1}}$. Reducing this value by 40 per cent, to account for non-denudational component of solute loads (see discussion above), the estimate is $2{,}200\,\mathrm{Mt\,a^{-1}}$ for denudational solute load. Globally, this is approximately 15 per cent of natural mechanical denudation. Chemical denudation rates although less variable than mechanical denudation rates, do vary significantly (Table 7). Reported values range from $1\,\mathrm{mm\,ka^{-1}}$ in drainage basins such as the Nile, Niger and Rio Grande to $27\,\mathrm{mm\,ka^{-1}}$ in the Chiang Jiang basin (Summerfield and Hulton 1994).

Some of the measured variability is related to lithology. Maximum yields of $6{,}000\,\mathrm{t\,km^{-2}\,a^{-1}}$ occur in rare instances (for example in areas underlain by halite). More usual maxima are $1{,}000\,\mathrm{t\,km^{-2}\,a^{-1}}$ in limestone regions. Although few studies have attempted to reconcile laboratory-based experimental studies of weathering rates and catchment scale estimates, the real-world weathering rates of different lithologies do correspond qualitatively to rates measured in the laboratory (Drever and Clow 1995). Many studies have documented a consistent positive correlation between solute load and annual runoff. This results from more water available for chemical reactions in the regolith and solute release, and greater runoff to transport these solutes. The relationship with temperature tends to be very weak (overwhelmed by other variables, especially precipitation and local relief). Relief influences a number of factors which impact the rate of surface runoff, rate of subsurface drainage and therefore rate of leaching of soluble constituents, and rate of erosion of weathered products and thereby rate of exposure of fresh mineral surfaces. In the Amazon basin, a relationship between relief and chemical weathering exists: ~86 per cent of the solutes delivered by the Amazon to the Atlantic come from the Andes mountains (~12 per cent of the area) (Gibbs 1967). The problem, however, is that for the Amazon relief and lithology are highly correlated; outcrops of limestone and evaporates are common in the Andes, whereas most of the remainder of the basin is underlain by silicate rocks. Based on data for externally draining basins exceeding $5 \times 10^5\,\mathrm{km^2}$ in area, Summerfield and Hulton (1994) conclude that chemical denudation rates are more strongly associated with relief than climatic factors. This supports the idea that the efficient removal of bedrock in the weathering zone is the critical determinant of the rate of chemical weathering.

Table 7 Solute denudational loads of major rivers in relation to climate and relief

Climate and relief zone	Denudational solute load ($\mathrm{t\,km^{-2}\,a^{-1}}$)	Total denudation ($\mathrm{mm\,ka^{-1}}$)	Typical solute load as % of total
Mountainous			
High precipitation	70–350	95–740	10
Low precipitation	10–60	45–370	10
Moderate relief			
Temperate or Tropical climate	25–60	30–110	35
Low relief			
Dry climate	3–10	5–35	10
Temperate climate	12–50	15–30	65
Subarctic climate	5–35	5–15	80
Tropical climate	2–15	1.5–10	50

Sources: Adapted from Summerfield (1991) based on Meybeck (1976)

Overall, high rates of chemical denudation are found in humid mountainous regions, where high relief is coupled with high runoff (Table 7). Minimum rates tend to be recorded in semi-arid regions where runoff is very low (although concentrations of dissolved load may be high), and in high latitude lowland terrains where runoff and solute concentrations are low. In some basins, especially those in a predominantly humid lowland environment, chemical denudation exceeds mechanical denudation. The other extreme are basins where extremely high sediment yields mean that chemical denudation represents less than 5 per cent of total denudation. Proportionally chemical denudation tends to become lower in drainage basins experiencing higher total denudation rates (Table 7).

Chemical denudation and global climate change

Given chemical denudation is an important control on the biogeochemistry of ecosystems, its study has implications not only for landform development but also global environmental change, notably issues of water quality, watershed acidification, nutrient cycling, and the greenhouse effect. As described above, chemical denudation is influenced by climate. Over geological time periods, however, chemical weathering also has a significant influence on global climate. During the weathering of carbonates and silicates, atmospheric CO_2 is taken up and converted to dissolved HCO_3^-. The HCO_3^- after delivery to the oceans by rivers can be stored in the form of carbonate minerals or organic matter in sediments. Either way there is a net loss of CO_2 from the atmosphere. Given CO_2 is a greenhouse gas, any changes in its concentration affects radiative exchanges in the Earth's atmosphere (Berner *et al.* 1987). By way of example, increased rates of chemical weathering associated with the Himalayan–Tibetan uplift have been suggested as a primary cause of the late Cenozoic ice ages (Raymo and Ruddiman 1992).

References

Berner, Robert A., Lasaga, Antonio G. and Garrels, Robert M. (1987) The Carbonate-Silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years, *American Journal of Science* 283, 641–683.

Drever, J.I. and Clow, D.W. (1995) Weathering rates in catchments, in A.F. White and S.L. Brantley (eds) *Chemical Weathering Rates of Silicate Minerals*, Mineralogical Society of America, *Reviews in Mineralogy* 1, 407–461.

Dole, R.B. and Stabler, H. (1909) *Denudation*, US Geological Survey Water Supply Paper 234, 78–93.

Gibbs, R.J. (1967) The geochemistry of the Amazon system I. The factors that control the salinity and the composition and concentration of the suspended solids, *Geological Society of America Bulletin* 78, 1,203–1,232.

Likens, G.E., Bormann, F.H., Pierce, R.S., Eaton, J.S. and Johnson, J.M. (1977) *Biogeochemistry of a Forested Ecosystem*, New York: Springer-Verlag.

Meybeck, M. (1976) Total dissolved transport by world major rivers, *Hydrological Sciences Bulletin* 21, 265–284.

Meybeck, M. (1987) Global chemical weathering of surficial rocks estimated from river dissolved loads, *American Journal of Science* 287, 401–428.
Raymo, M.E. and Ruddiman, W.F. (1992) Tectonic forcing of late Cenozoic climate, *Nature* 359, 117–122.
Riebe, C.S., Kirchner, J.W., Granger, D.E. and Finkel, R.C. (2001) Strong tectonic and weak climatic control on long-term chemical weathering rates, *Geology* 29, 511–514.
Summerfield, M.A. (1991) *Global Geomorphology*, Harlow, England: Longman.
Summerfield, M.A. and Hulton, N.J. (1994) Natural controls of fluvial denudation rates in major world drainage basins, *Journal of Geophysical Research – Solid Earth* 99, 13,871–13,883.
White, A.F. (1995) Chemical weathering rates of silicate minerals in soils, in A.F. White and S.L. Brantley (eds) *Chemical Weathering Rates of Silicate Minerals*, Mineralogical Society of America, *Reviews in Mineralogy* 1, 407–461.

SEE ALSO: weathering; weathering and climate change

CATHERINE SOUCH

CHEMICAL WEATHERING

The biogeochemical alteration of the Earth's surface and associated processes are called WEATHERING. These processes are usually separated into chemical, physical and biologic weathering for discussion. In reality, these processes are not mutually exclusive. Chemical weathering is the process by which chemical reactions such as hydrolysis, hydration, oxidation-reduction, ion exchange, solution and organic reactions transform rocks and minerals into new chemical combinations that are stable under conditions at or near the Earth's surface. Chemical weathering begins as thermodynamically unstable minerals adjust to the surrounding environment. Rocks and minerals that are not in equilibrium with near-surface conditions of temperature, pressure and water begin to alter to new products that are chemically more stable in the near surface.

Chemical weathering processes are many. The ability to measure these processes has progressed over the years as newer technologies and interdisciplinary research have led to discoveries at all scales from the molecular to the macroscale. Although chemical weathering occurs at many different temperatures and pressures this discussion will focus on a few basic concepts common to weathering under near-surface conditions.

The resistance of rocks and minerals to chemical breakdown influences the stability of individual mineral species in the environment. This stability is related to several mineral properties including cleavage and fracture patterns, particle size and specific surface, solubility and the relative stability of the surrounding environment. Structurally, the resistance to weathering increases as the complexity of silicate linkage increases, particularly the number of shared oxygens, from independent tetrahedral structures (e.g. olivine) to single chain silicates (e.g. enstatite, a pyroxene) to sheet silicates (e.g. talc) to continuous framework silicates (e.g. quartz). A weathering sequence that illustrates this concept for common rock-forming silicate minerals is illustrated in Figure 23. Stability of minerals increases from top to bottom. Additional guides to mineral stability are discussed in Ritter (1986).

Other factors being equal, minerals formed in environments resembling those in which weathering takes place will be the most resistant. This concept is based on thermodynamic principles. For example, olivines and calcium plagioclase feldspars form at higher temperatures and pressures and weather more rapidly than muscovite and quartz-rich minerals which form at lower temperatures. These latter conditions are more similar to near-surface weathering conditions.

Chemical weathering processes

Solution occurs when a mineral dissolves to form ions or dispersed colloidal molecular units. It is one of the simplest of weathering processes. Bicarbonate ($2HCO_3^-$) is derived from the dissociation of carbonic acid (H_2CO_3) that in turn formed from the dissolution of carbonate rock and atmospheric CO_2 dissolved in water:

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

Bicarbonate is one of the most abundant anions in weathering systems. Its weathering effects have been studied in detail relative to limestone KARST systems. Bicarbonate ions can also form from dissolution of CO_2 in plant and microbial respiration processes:

$$H_2O + CO_2 \leftrightarrow H_2CO_3 \leftrightarrow H^+ + HCO_3$$

HYDROLYSIS is the reaction of compounds with water to produce a weak acid or weak base. Water molecules are attracted to surfaces of minerals due to the attraction of polar water molecules to the polar surfaces of many minerals. Here, forsterite hydrolysis produces silicic acid:

$$Mg_2SiO_4\ \text{(forsterite)} + 4H_2O \leftrightarrow 2Mg^{++} + 4OH^- + H_4SiO_4\ \text{(silicic acid)}$$

Natural waters usually contain dissolved CO_2 so reactions often contain carbonic acid as well. A more complete way to write the above reaction in a natural system would be:

$$Mg_2SiO_4\ \text{(forsterite)} + 4H_2CO_3 \leftrightarrow 2Mg^{++} + 4HCO_3^- + H_4SiO_4$$

H^+ (from water) can replace other ions such as K^+, Ca^{++} and Na^+ in mineral structures. The H^+ disrupts the structural bonds. If the H^+ is smaller than the cation it replaces, physical strain occurs in the mineral which in turn accelerates weathering. For example, microcline feldspar reacts with water and loses a potassium ion:

$$KAlSi_3O_8\ \text{(microcline feldspar)} + H_2O \rightarrow HAlSi_3O_8 + K^+ + (OH)^-$$

Lowering the pH increases hydrolysis because the number of H^+ ions in solution increases. For example, organic matter decomposition adds H^+ and speeds hydrolysis as do many other biologic processes such as nutrient uptake, nitrification and sulphur oxidation. Warm temperatures have an effect similar to lowered pH. Higher temperatures increase the dissociation of water molecules and provide additional H^+, potentially increasing hydrolysis in a system. Thus the microcline feldspar in the above example should weather more quickly in a warmer rather than a cold climate and in an acid rather than a more neutral environment.

HYDRATION adds water molecules to mineral structures but the water does not dissociate as in hydrolysis. Gypsum is a hydrated form of anhydrite. The reverse reaction is dehydration:

$$CaSO_4\ \text{(anhydrite)} + 2H_2O \leftrightarrow CaSO_4 \cdot 2H_2O\ \text{(gypsum)}$$

Although quartz is a resistant mineral, under specific conditions it can dissolve by hydration:

$$SiO_2\ \text{(quartz)} + 2H_2O \leftrightarrow H_4SiO_4$$

Some minerals may expand during hydration. Commonly smectite hydrates and dehydrates

Olivine (Mg, Fe)
\
Augite (Ca, Mg, Fe, Al)
\
Hornblende (Ca, Na, Mg, Fe, Al)
\
Biotite (K, Mg, Fe, Al)
\

Calcic plagioclase (Ca, Al)
/
Calcic-alkali plagioclase (Ca, Na, Al)
/
Alkali-calcic plagioclase (Na, Ca, Al)
/
Alkali plagioclase (Na, Al)
/

Potash feldspar (K, Al)
|
Muscovite (K, Al)
|
Quartz (Si)

Figure 23 Weathering of common rock-forming silicate minerals
Source: Data from Goldich, S.S. (1938) A study in rock weathering, *Journal of Geology*, 46, 17–58

when water molecules enter or leave interlayers, respectively. In an expanded condition, minerals are more porous and become more susceptible to additional weathering.

Ion-exchange reactions are important and are usually related directly to clay mineral weathering and other secondary minerals because these minerals have a high capacity for exchange within the interlayer and with surface ions. During exchange, the basic structure of the mineral is unchanged, but interlayer spacing varies with each cation absorbed into the interlayer. This mechanism has a unique outcome for clay minerals in that the alteration of one clay mineral may produce another. For example, under certain circumstances smectite may form from illite with the loss of interlayer K^+. Ion exchange is an important factor in biogeochemical reactions of rocks and sediments with organic matter and colloids. Ion exchange can also occur in the initial weathering of primary minerals such as silicates.

Ion mobility is key to primary mineral weathering. Hudson (1995) discusses an update of Polynov's 1937 ion mobility series that ranks major elements from very mobile (I) to relatively immobile (V):

$$Cl > SO_4 > Na > Ca > Mg > K > Si > Fe > Al$$

where mobility phase I is Cl and SO_4, II is Na, III is Ca, Mg and K, IV is Si and V is Fe and Al. Mobility depends on charge and charge density.

In a strongly leaching environment only phase V elements would remain. As an environment became drier, phase IV through I elements would become increasingly abundant. For example, gibbsite ($Al[OH]_3$) is a common aluminium hydroxide in sediments and soils assumed to be in the latter stages of weathering where leaching conditions and free drainage occur. This would be equivalent to phase V ion mobility. At this point, silica has been so thoroughly removed from the system that phyllosilicates can no longer form. Aluminium hydroxide-rich sediments are associated with tropical environments today and in the weathering profiles of bauxite deposits of ancient silica-depleted rock systems.

OXIDATION and REDUCTION equilibria, also known as redox reactions, take place when an atom or element gains or loses net charge; oxidation, the loss of electrons and reduction, the gaining of electrons. The availability or absence of one electron acceptor leads to the reduction of another element. Elements must have at least two viable oxidation states to be involved in redox reactions. Only about six elements, oxygen, iron, manganese, sulphur, nitrogen and carbon, are abundant enough in the natural environment to take part in common redox reactions of the near-surface environment. Oxygen plays a role in most oxidation processes. In the reaction below, ferrous iron derived from the hydrolysis of an iron-bearing silicate, is oxidized from +2 to +3 oxidation state to form haematite. Oxygen is reduced from 0 to −2:

$$2Fe^{++} + 4HCO_3^- + \tfrac{1}{2}O_2 \leftrightarrow Fe_2O_3 \text{ (haematite)} + 4CO_2 + 2H_2O$$

Haematite is stable in many environments but goethite, also a ferric iron component and primary constituent of limonite, may occur with the addition of more moisture:

$$Fe_2O_3 \text{ (haematite)} + H_2O \leftrightarrow 2HFeO_3 \text{ (goethite)}$$

Oxidation of pyrite, FeS_2 to iron hydroxides or sulphates and sulphuric acid on exposure to water and oxygen has detrimental consequences. This reaction, often occurring in materials adjacent to mine sites, is a common cause for the sterile biologic conditions in sediments drained by acid waters. The term acid mine drainage is applied to these waters in which the pH can drop below 2.

Oxidation of organic carbon is often due to micro-organisms that play a major role in expediting redox reactions. An example of an organic oxidation reaction in which carbon dioxide is formed is:

$$C_6H_{12}O_6 + 6O_2 \leftrightarrow 6CO_2 + 6H_2O$$

The carbon dioxide formed is available for solution and hydrolysis reactions.

Chelation (see CHELATION AND CHELUVIATION), a form of metal complexation, is the reaction between a metallic ion and a complexing agent, usually organic, resulting in the formation of a ring structure that encompasses the metallic ion effectively removing it from the system. Hydrogen is often released during the process and becomes available for hydrolysis reactions. Chelating agents in contact with rocks or minerals can cause significant weathering (Berthelin 1988). For example, lichens and mosses remove cations from silicate minerals and may produce dissolved or amorphous silica. Some breakdown of minerals occurs from reactions with organic acids produced at the root tips of plants or produced by bacteria acting on decaying material.

Chemical weathering products

Chemical weathering results in either congruent dissolution, in which the material goes completely into solution, or incongruent dissolution, in which at least some weathering products may form new minerals (neoformation or synthesis) or leave a residue or precipitate. If limestone dissolves completely and releases Ca^{++} and HCO_3^- ions into aqueous solution it is a congruent dissolution. However, most limestones are not pure $CaCO_3$ and leave a residue. During chemical changes, particle size decreases, surface area increases and constituents continue to dissolve into aqueous weathering solutions. Water is often the transferring agent and its activity is important.

Berner (1971) and Berner and Berner (1996) emphasize the importance of water flow as a control factor on the intensity of weathering. Berners' example suggests that at moderate flow rates albite alters to kaolinite but at higher flow rates, silicic acid is removed so quickly that gibbsite rather than kaolinite may form. When flow rates were very slow, material was removed slowly and if magnesium was available, the product was montmorillonite. This suggests that climate and relief control weathering products. The mineralogy of the rock weathering and the chemical composition of weathering solutions are two additional determining factors. Chadwick *et al.* (2003) present a biogeochemical model for an arid to humid climosequence on Kohala Mt., Hawaii. They found that where mean annual precipitation is high and total sediment pore space is annually full, leaching of soluble base cations and silica is nearly complete. At lower precipitation inputs, leaching losses are progressively lower. Secondary mineral weathering was controlled by metastable non-crystalline weathering products rather than soil solution composition.

Weathering products may be grouped into four categories: (1) soluble constituents; (2) residual primary minerals unaffected by weathering reactions; (3) new stable minerals produced by weathering reactions; (4) organic compounds. Soluble constituents are those that remain in solution at near-surface conditions. Three primary groups of residual minerals remain in weathered soils: (a) phyllosilicate clay minerals; (b) very resistant end products such as sesquioxides of Fe and Al; (c) very resistant primary minerals such as quartz, zircon and rutile. Each group contributes less as weathering progresses. In highly weathered soils and sediments of the humid tropics or subtropics, Al and Fe oxides and low-activity clay minerals with low Si/Al ratios may be all that remains of the original primary minerals. Feldspars, mica, amphiboles and pyroxene minerals alter to clay minerals through hydrolysis, hydration and oxidation. For example, biotite mica weathers as Fe^{++} oxidizes, K^+ leaves the structure to maintain neutrality, the structure begins to weaken and soluble cations in solution such as Ca^{++}, Mg^{++} or Na^+ replace the remaining K^+. A new phyllosilicate such as vermiculite or montmorillonite forms.

Phyllosilicates are commonly stable mineral products of weathering. They are specific clay minerals occurring primarily in the clay-size fraction of a material (see Moore and Reynolds 1997). Phyllosilicates strongly influence the chemical as well as physical properties of sediments, in part due to their unusually small particle size and resulting high surface area but also, as described earlier, due to cation exchange characteristics uniquely related to their crystal structures (see Dixon and Weed 1989; Moore and Reynolds 1997). Linus Pauling (1929, 1930), Kelley (1948) and Grim (1962) were some of the first individuals to recognize the unique chemical properties of phyllosilicate clay minerals.

Chemical weathering and landscapes

Measurement of the total amount of chemical weathering is important to a geologist or geomorphologist because it can provide some estimate of landscape evolution. Although both physical and chemical denudation affects landscapes on a catchment or global scale this discussion centres on chemical weathering. Chemical denudation can be calculated from dissolved stream loads and corrected for atmospheric input because most ions in water come from weathering reactions (Berner and Berner 1987). Annual load is multiplied by annual discharge and divided by basin area. Berner and Berner (1987) have calculated a world average. Garrels and MacKenzie (1971) ranked chemical denudation by continent: Europe > North America = Asia > South America > Africa ≫ Australia. Degree of weathering is often calculated on the basis of total chemical analyses comparing fresh parent rock with saprolite or soil derived *in situ* from it. Birkeland (1999) presents a good summary of this approach.

During physical weathering in an open system, the landscape is generally lowered volumetrically because solids are removed but during chemical weathering the landscape may increase volumetrically. Ions removed from a weathering rock mass might be reflected in a bulk density change such that a geomorphic surface is unchanged or is even raised. For example, when a soil forms from rock, the bulk density will decrease, sometimes by 0.5 $g\,cm^{-3}$ or more. This results in an overall volumetric expansion (Birkeland 1999). Brimhall and others (1991) developed a method for assessing chemical change during weathering that gives values for volume change as well as for losses or gains in mass. In some cases this expansion is the catalyst for increased physical weathering. Several researchers have suggested that the formation of grus from granite follows these steps (e.g. Wahrhaftig 1965; Nettleton *et al.* 1970).

References

Berner, E.K. and Berner, R.A. (1987) *The Global Water Cycle; Geochemistry and Environment*, Englewood Cliffs, NJ: Prentice Hall.

Berner, E.K. and Berner, R.A. (1996) *Global Environment: Water, Air, and Geochemical Cycles*, Englewood Cliffs, NJ: Prentice Hall.

Berner, R.A. (1971) *Principles of Chemical Sedimentology*, New York: McGraw-Hill.

Berthelin, J. (1988) Microbial weathering processes in natural environments, in A. Lerman and M. Meybeck (eds) *Physical and Chemical Weathering in Geochemical Cycles*, 33–59, The Netherlands: Kluwer.

Birkeland, P.W. (1999) *Soils and Geomorphology*, New York: Oxford University Press.

Brimhall, G.H., Chadwick, O.A., Lewis, C.J., Compston, W., Williams, I.S., Danti, K.J., Dietrich, W.E., Power, M.E., Hendricks, D. and Bratt, J. (1991) Deformational mass transport and invasive processes in soil evolution, *Science* 255, 695–702.

Chadwick, O.A., Gavenda, R.T., Kelly, E.F., Ziegler, K., Olson, C.G., Elliot, W.C. and Hendricks, D.M. (2003) The impact of climate on the biogeochemical functioning of volcanic soils, *Chemical Geology*, in press.

Dixon, J.B. and Weed, S.B. (eds) (1989) *Minerals in Soil Environments*, 2nd edition, Soil Science Society of America Book Series 1, Madison: Soil Science Society of America.

Garrels, R.M. and MacKenzie, F.T. (1971) *Evolution of Sedimentary Rocks*, New York: W.W. Norton.

Goldich, S.S. (1938) A study in rock weathering, *Journal of Geology* 46.

Grim, R.E. (1962) *Applied Clay Mineralogy*, New York: McGraw-Hill.

Hudson, B.D. (1995) Reassessment of Polynov's ion mobility series, *Soil Science Society of America Journal* 59, 1,101–1,103.

Kelley, W.P. (1948) *Cation Exchange in Soils*, America Chemical Society Monograph, New York: Reinhold.

Moore, D.M. and Reynolds, R.C. Jr. (1997) *X-ray Diffraction and the Identification and Analyses of Clay Minerals*, New York: Oxford University Press.

Nettleton, W.D., Flach, K.W. and Nelson, R.E. (1970) Pedogenic weathering of tonalite in southern California, *Geoderma* 4, 387–402.

Pauling, L. (1929) The principles determining the structure of complex ionic crystals, *Journal of the American Chemical Society* 51, 1,010–1,026.

——(1930) The structure of micas and related minerals, *Proceedings National Academy of Science USA* 16, 123–129.

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

Wahrhaftig, C. (1965) Stepped topography of the southern Sierra Nevada, California, *Geological Society of America Bulletin* 76, 1,165–1,190.

Further reading

Bartlett, R.J. and James, B.R. (1993) Redox chemistry of soils, *Advances in Agronomy* 50, 151–208.

Brindley, G.W. and Brown, G. (eds) (1980) *Crystal Structures of Clay Minerals and their X-ray Identification*, Monograph No. 5, London: Mineralogical Society.

Brownlow, A.H. (1979) *Geochemistry*, Englewood Cliffs, NJ: Prentice Hall.

Clayton, J.L., Megahan, W.F. and Hampton, D. (1979) *Soil and Bedrock Properties: Weathering and Alteration Products and Processes in the Idaho Batholith*, USDA Forest Service Research Paper INT-237.

Colman, S.M. and Dethier, D.P. (eds) (1986) *Rates of Chemical Weathering of Rocks and Minerals*, Orlando, FL: Academic Press.

Drever, J.I. (1982) *The Geochemistry of Natural Waters*, Englewood Cliffs, NJ: Prentice Hall.

——(ed) (1985) *The Chemistry of Weathering*, NATO-ASI, The Netherlands: Reidel Publishing.

——(1997) Weathering processes, in O.M. Saether and P. de Caritat (eds) *Geochemical Processes, Weathering and Groundwater Recharge in Catchments*, 3–19, Rotterdam: A.A. Balkema.

Garrels, R.M. and Christ, C.M. (1965) *Solutions, Minerals and Equilibria*, New York: Freeman and Cooper.

Greenland, D.J. and Hayes, M.H.B. (1983) *The Chemistry of Soil Constitutents*, New York: John Wiley.

Helling, C.S., Chester, G. and Corey, R.B. (1964) Contribution of organic matter and clay to soil cation-exchange capacity as affected by the pH of the saturation solution, *Soil Science Society of America Proceedings* 28, 517–520.

James, B.R. and Bartlett, R.J. (2000) Redox phenomena, in M.E. Sumner (ed.) *Handbook of Soil Science*, B169–B194, Boca Raton, FL: CRC Press.

Krauskopf, K.B. (1979) *Introduction to Geochemistry*, 2nd edition, New York: McGraw-Hill.

Newman, A.C.D. (ed.) (1987) *Chemistry of Clays and Clay Minerals*, Mineralogical Society of England, Monograph No. 6, Harlow, Essex: Longman.

Sparks, D.L. (1995) *Environmental Soil Chemistry*, New York: Academic Press.

Thurman, E.M. (1985) *Organic Geochemistry of Natural Waters*, Hingham, MA: Kluwer.

White, A.F. and Brantley, S.L. (eds) (1995) Chemical weathering rates of silicate minerals, *Reviews in Mineralogy* 31, 583.

Yuan, T.L., Gammon, N. Jr. and Leighty, R.G. (1967) Relative contribution of organic and clay fractions to cation-exchange capacity of sandy soils from several groups, *Soil Science* 104, 123–128.

SEE ALSO: dissolution; leaching; solubility; weathering

CAROLYN G. OLSON

CHENIER RIDGE

Chenier ridges (cheniers) are sandy or shelly elongate BEACH RIDGES, differentiated from other sand or shell beach ridges by the fact that they are perched on and separated laterally from other cheniers on a chenier plain, by fine-grained, muddy (or sometimes marshy) sediments. Other types of barrier beach plains can be mistaken for cheniers if the presence of underlying and interspersing muddy sediments is not adequately determined (normally by coring). Chenier ridges frequently bend landward at the downdrift end, and branch in a fan-like fashion. The name derives from the French word chêne, meaning oak, which grows on the Louisiana USA chenier ridges. Cheniers can be up to 6 m high, tens of kilometres in length, and hundreds of metres wide. Chenier plains can be tens of kilometres wide. Cheniers are found on generally low wave energy, low gradient, muddy shorelines, in areas where there is an abundant sediment supply. They are frequently associated with river deltas and bayhead situations. Although reported at high latitudes, most examples occur in tropical or subtropical locations. Augustinus (1989) provides an overview of examples and presumed examples of cheniers. Among the most reported examples are: the west Louisiana and Texas coast; Suriname, Guyana and French Guiana; the Gulf of California; New Zealand; northern Australia; east China.

Local variations in sediment supply (such as periods of different river discharge) have been suggested as the likely cause of alternate mudflat progradation and chenier ridge deposition (Otvos and Price 1979), although synchronous development of mudflat and chenier ridges has also been reported (Woodroffe *et al.* 1983; Woodroffe and Grime 1999). Periods of higher wave energy, however, are generally regarded as providing the means by which coarser sediments (including shells) are winnowed out for accumulation in the chenier ridge, with these sediments then moved landward by wave action and OVERWASHING. Some authors argue that 'true cheniers' must result from transgressive processes; however, Otvos (2000) believes the term is appropriate for both stranded (regressive) and transgressive cheniers (see TRANSGRESSION). Otvos (2000) also argues that beach ridges fronting chenier plains must become isolated from the sea and inactive by the deposition of mudflats on their seaward side before they can be considered as cheniers. Chenier plains can provide a sensitive record of changes in sediment supply, sea level and environmental processes.

References

Augustinus, P.G.E.F. (1989) Cheniers and chenier plains: a general introduction, *Marine Geology* 90, 219–229.

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

Otvos, E.G. and Price, W.A. (1979) Problems of chenier genesis and terminology – an overview, *Marine Geology* 31, 251–263.

Woodroffe, C.D. and Grime, D. (1999) Storm impact and evolution of a mangrove-fringed chenier plain, Shoal Bay, Darwin, Australia, *Marine Geology* 159, 303–321.

Woodroffe, C.D, Curtis, R.J. and McLean, R.F. (1983) Development of a chenier plain, Firth of Thanes, New Zealand, *Marine Geology* 53, 1–22.

SEE ALSO: beach ridge; overwashing; raised beach; transgression

KEVIN PARNELL

CHRONOSEQUENCE

The term is used to describe a series of soils that reflect the importance of time for soil formation. *Inter alia*, young soils will differ from mature soils in the degree of weathering of the soil parent material, the development of the soil horizons and the abundance of secondary minerals.

Because the time spans involved in soil development are beyond the time frame of direct observation, usually the development of soils of different age are compared. A chronosequence is thus a sequence of related soils that differ from one another in certain properties primarily as a result of the time available for soil formation. Classical examples are the soils developed on the different members of a flight of terraces, where – except for time – all soil-forming factors (as parent material, landform, climate, etc.) should be rather similar.

Different types of chronosequences can be distinguished: (1) post-incisive, (2) pre-incisive, and (3) time-transgressive. The most frequently studied is case (1) – the example mentioned above – where soils evolve on a sequence of surfaces of different age. In (2) soils that began to develop on a particular surface at the same time, but that were subsequently buried at different times at different places, form a chronosequence. Case (3) relates to a vertical stacking of sediments and PALAEOSOLS, i.e. soils that formed on the same place, but that have been buried after differing periods of development.

Chronosequences have been used to establish quantitative descriptions of soil changes with time, called chronofunctions, and to use the degree of soil development for estimating soil age. To allow for quantitative estimates, soils on dated surfaces (see DATING METHODS) are investigated and numerical indices, such as eluvial-illuvial coefficients and soil development indices, have been developed. There are limits to the range over which chronofunctions can be applied. The rates of development of most soil properties decrease with time; once this degree of development has been achieved, further inferences regarding time cannot be made. In addition, many more complex functions with clear thresholds are involved in soil development. Establishment of chronosequences is difficult in many cases because with the passage of time other soil-forming factors usually also change – as is most clear in the case of climate (see PALAEOCLIMATE). It is often also difficult to rule out the influence of soil disturbances and soil erosion.

Chronosequences have played an important role in establishing relative soil chronologies, which in turn have been used to establish stratigraphic relationships for different geomorphic surfaces. This is especially important where due to the lack of suitable methods or materials modern chronometric dating techniques cannot be applied.

Further reading

Birkeland, W.P. (1999) *Soils and Geomorphology*, 3rd edition, New York: Oxford University Press.

SEE ALSO: catena; soil geomorphology; weathering

ANDREAS LANG

CIRQUE, GLACIAL

Definition and Form

Cirques, also known as corries, coves, combes or cwms, are hollows formed at glacier sources in mountains and partly enclosed by steep, arcuate slopes (headwalls) (Plate 22). Cirque formation requires deepening of the floor by glacial plucking and abrasion, plus glacial removal of plucked or fallen rock encouraging continued headwall retreat. These are aided by basal slip and rotational flow of steep glaciers.

A well-developed 'armchair cirque' has a gently sloping floor and a steep headwall (giving profile closure). At least some of the floor should be gentler than 20°. The headwall curves around the floor, giving plan closure. Ideally, the floor ends in a distinct threshold beyond which the slope steepens, but this may be absent in a trough-head cirque. The headwall should exceed the angle of talus (about 31°–35°) at least in part. We can draw the boundary between headwall and floor at an angle of some 27° (a 2 mm spacing of 10 m contours on a 1 : 10,000 map). A similar gradient can be used to define the cirque crest at the top of the headwall if there is a gentler slope above. It is useful to define a 'cirque focus' in the middle of the threshold. A line from there to the top of the headwall, dividing the cirque into two halves equal in map area, to left and to right, is the median axis: this is used to measure length and overall aspect (Evans and Cox 1995).

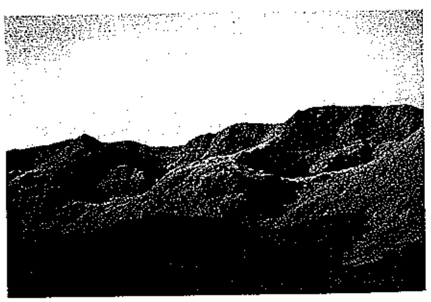

Plate 22 East-facing cirques in Ordovician volcanic tuffs on the ridge south of Helvellyn, English Lake District; from left to right, Cock and Ruthwaite Coves, Hard Tarn (a smaller hollow) and Nethermost Cove. The cirques hang above Grisedale trough

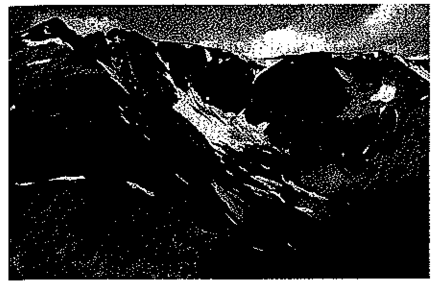

Plate 23 North-facing cirque in Triassic metamorphic rocks on Mount Noel (2,600 m), south of Bralorne in the Coast Mountains of British Columbia. The small glacier is a remnant of a larger Little Ice Age cirque glacier which formed the two sharp lateral moraines at bottom right (August 2000)

Cirque size varies over an order of magnitude, and provides a characteristic scale to glaciated mountains: cirques are scale-specific landforms, averaging around 700 m long and broad, and a few hundred metres deep. Overall centre-line gradients, approximating those of glaciers filling the cirques, vary from 5° to 50°, with means commonly 20° to 25° (length/depth 2.1 to 2.8).

Cirque form is simple in plateau areas, more complicated in high-relief areas of coalescing cirques, and very difficult to define where modified by overriding ice sheets. Valley-head and valley-side cirques are often distinguished, but a more important distinction is between 'armchair cirques', the type normally described, and 'high-alpine cirques' found in high massifs of the European Alps and in coastal regions of British Columbia, Washington and Alaska (Plate 23). 'High-alpine' cirques are shallow and have steep, straight, abraded apron-like floors (20°–31°) hanging above troughs along which ice was rapidly evacuated. Classic armchair cirques are deeply concave in both profile and plan (concave contours). For both types, the concave break in slope between headwall and floor prevents a good fit to simple equations.

Processes

Ideas on processes of cirque development have changed considerably over time. The importance of glacial erosion was clearly established in the 1870s by Gastaldi and Helland for cirques in the Alps and Norway. Although glacial protectionists argued for fluvial or tectonic origins into the twentieth century, the existence of deep, rounded rock basins in many cirques could not be explained except by glacial erosion. Nevertheless it was accepted that processes acting around and under snowpatches (NIVATION) widened initial hollows before glaciers could become established: cirque widening formed the basis of a 'cycle of mountain glaciation' proposed by W.H. Hobbs. After 1906, W.D. Johnson's observation that frost action was active in and above the BERGSCHRUND – the initial crevasse as the glacier accelerates away from a cirque headwall – was accepted as an essential process of cirque development. Although cirques were used as evidence of former glaciation and in the reconstruction of former snowlines, their development by essentially periglacial processes was emphasized throughout the first half of the twentieth century.

Neither nivation nor frost weathering, however, explain the deepening of cirques by erosion of their floors. After a brief flirtation in the 1940s with the hypothesis of extrusion flow – the unrealistic idea that soft basal ice would be squeezed forward by the weight of overlying ice, without carrying the overlying ice forward – geomorphologists found a better way of obtaining fairly high basal ice velocities. Observations in the Jotunheim (Norway) in the late 1940s showed that steep glaciers banked against cliffs acted rather like landslides, and moved over their beds by rotational slip (McCall, and Grove; in Lewis 1960). In this way the glacial abrasion and plucking advocated by Helland became easier to explain. This also implied that basal ice must be wet, i.e. 'warm' – at its melting point.

Rotational slip, basal abrasion and plucking, and frost weathering around the bergschrund or upper glacier margin do not, however, cover the whole story of cirque development. Observations by Battle showed that temperatures within bergschrunds varied too slowly for the daily or seasonal frost cycle to be effective. Gardner (1987) showed that the *randkluft* or *rimaye* – the upper margin of a glacier against a cliff – is a more likely site for frost action, and migrates up and down the headwall over time. Whalley has pointed out that the stress field in high cliffs undercut by glaciers causes instability. With or without the help of frost action, stress concentration and release cause headwall collapse by rockfall and rock avalanche (see STURZSTROM). Many rock avalanches, both historic and older, are from cirque headwalls (Evans 1997). Active rockfall onto cirque glaciers can be observed today, especially as glacier wastage has increased the area of exposed headwall above.

While this mechanism helps to account for headwall retreat, fuller understanding of subglacial erosion has followed study of water pressure variations in glacier boreholes. Hooke (1991) and Iverson (1991) have shown that these are frequent and of high magnitude where crevasses (including the bergschrund) permit meltwater to reach the bed. On steep lee (down-ice) slopes, water-filled cavities tend to open. When water pressure falls, there is a delay before pressure falls in cracks in the rock: this aids crack extension. When water pressure rises, ice velocity increases and it is easier for joint-bounded blocks to be carried forward (entrained) in the basal ice. These neatly inter-related mechanisms provide what may well be the most important process for glacial plucking (quarrying).

Much more is now known about the frequent and large variations of climate during the Quaternary. This makes slow development with a snowpatch or glacier of a given size unlikely: a nivation phase is soon overtaken by a glacial phase. Wet-based glaciers erode much more rapidly than snowpatches. Erosion can be effective even where the margins of a glacier are frozen to the bed; Bennett *et al.* (1999) have described very rapid erosion by thrusting of sedimentary bedrock stressed by the transition from a sliding to a frozen subglacial boundary. Such 'polythermal' glaciers are common today in Svalbard, and in Sweden (Richardson and Holmlund 1996), and may deepen cirque floors as the warm, deep central ice slides over its bed. Adjacent rock may break down by growth and thawing of ice lenses as the zero isotherm migrates over the long term.

Cirque erosion thus requires basally sliding ice to quarry and abrade the bed, steepening the margins until they collapse. Erosion is aided by concentration of snow accumulation high on the glacier (Figure 24), and by gradients between 12° and 26°, both of which encourage rotation; also by rapid variations in basal water pressure. Steepened headwalls collapse by rockfall or rock avalanche, often on deglaciation but sometimes earlier or later.

Cirque development

Mountain glaciers form in concavities, deepening, widening and simplifying these to form glacial cirques. Starting with glacial occupation of any large hollow, their floors are deepened, headwalls retreat and concavity is increased. Initial concavities include gully heads, gully junctions, landslide scars, structural benches and volcanic craters. Diversity of initial concavities provides a broad range of poorly developed cirques: continued glacial erosion: (1) increases headwall gradient, (2) reduces floor gradient, (3) increases length, width and depth (amplitude), and (4) increases plan closure by eroding more deeply into the mountain mass, so that the headwall curves around the floor.

Rotational slip, deepening the floor, is favoured in cirque glaciers, but erosion continues even if glaciers extend beyond cirques. Gordon (1977) proposed a model of cirque development whereby cirques lengthen, broaden, deepen and increase their concavity as they enlarge. Nevertheless, correlations between the measures of development (headwall gradient, inverse floor gradient, and plan closure) are very weak, implying that cirques develop along varied paths and the influence of initial site remains important. Thus cirque form is very diverse.

Larger cirques are better developed on all criteria: they are also flatter, i.e. horizontal dimensions increase more rapidly than vertical. This has been interpreted as allometric development (see ALLOMETRY), although it relates to spatial rather than temporal variation (Olyphant 1981). A peak or a high headwall on the equatorward side of a cirque helps preserve snow: a pass to windward funnels more snow in (Graf 1976). Better developed cirques are thus more effective in sheltering glaciers. This gives a positive feedback, so that a steady state is unlikely unless surrounding ridges are lowered at the same rate as all parts of the cirque.

Where initial hollows are closely spaced, the formation and retreat of cliffed headwalls eventually sharpen the intervening ridge into an ARÊTE. Gullies and structural irregularities in cliffs lead to marked rises and falls in arête crests, giving a series of GENDARMES. Since cirque glaciation is commonly asymmetric, arêtes arise from lateral intersection, i.e. two sidewalls retreating into each other. When the EQUILIBRIUM LINE OF GLACIERS falls sufficiently for glaciers to form on opposing slopes, e.g. east and west, steeper arêtes flank a deep parabolic col of 'interosculation'.

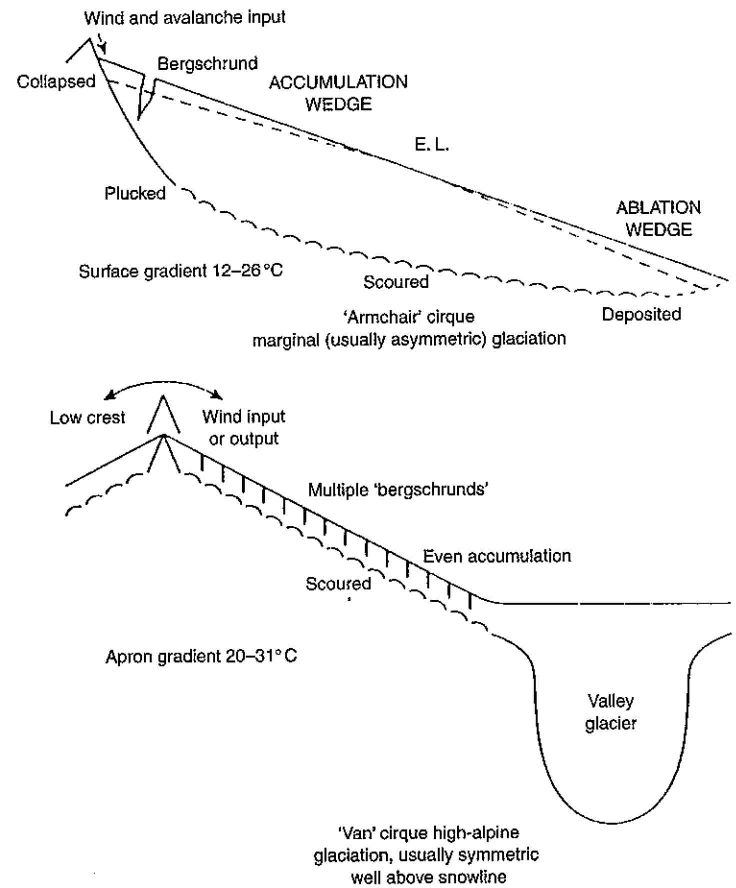

Figure 24 A model of the contrast between classical armchair cirques (above) and high-alpine cirques (from Evans 1997: 162)

Lateral enlargement continues, perhaps by coalescence, with no clear upper limit so long as a mountain mass remains. In the Antarctic, troughs and cirques have developed over a much longer period of glaciation and some cirque complexes are very large: forms of local glaciation developed in the earlier part of this period, and many are now 'fossilized' under very cold ice.

Many areas have suffered both local and ice sheet glaciation and cirques away from ice divides have been further modified (degraded) by overriding ice. On deglaciation, cirques are degraded by subaerial mass movements, talus accumulation and gullying: these reduce headwall gradient and obscure the floor, but do not affect plan closure.

Relation to geology, climate and topography

Cirques form on all rock types, but postglacial degradation of headwalls is most likely on the weakest rocks such as shales. The expected effect of rock liability to abrasion has not been demonstrated, but the importance of joint sets is often noted (Haynes 1968). Inward-dipping joints or beds favour excavation of a rock basin, and these are more common on crystalline rocks. Simple, rounded cirques are found on homogeneous or frequently alternating rocks, e.g. flat-lying volcanic and sedimentary rocks. Greatest distortions in form come from single major contrasts, for example, juxtaposition of limestone and shale or quartzite and gneiss, giving a floor or a steeper side on the less-jointed rock. Cirque headwall heights and gradients should relate to ROCK MASS STRENGTH: results to date, however, show considerable scatter.

Well-developed cirque floors relate to the former snowline (Equilibrium Line), for example in being much lower on windward sides of major mountain ranges (Derbyshire and Evans 1976). This means they relate to snowfall rather than to the freezing level. More locally, snow is blown to leeward slopes and preserved longer on shady slopes, and cirques are more frequent on corresponding aspects.

In some regions, cirques are separated from each other by plateau areas or rolling topography: elsewhere they intersect and form more arêtes and horns. In the early twentieth century, influenced by the Davisian model, this was regarded as a developmental sequence. It is more likely that these contrasts relate to regional topography (Gordon 2001), including relief and drainage density, due ultimately to tectonic setting and climatic history.

Further work

Most morphometric studies have been confined to single regions, or based on selected cirques: consistent results from complete populations of cirques are needed, to establish variations between different regions and to start accounting for variations in relation to climate, geology and topography.

We have little information on periods of time required for cirque development, though currently observed rates of glacial erosion are adequate to erode cirques in a few hundred thousand years – only a proportion of the Quaternary. Headwall retreat per glaciation would be some 10 m, and floor deepening a few metres. New techniques of exposure dating of surfaces only tell us when ice last disappeared; techniques such as fission-track dating give some idea of rock uplift history, but with very broad error margins. More precise dating approaches are needed to provide specific chronologies of cirque development. In small regions cirques may have developed simultaneously (Evans 1999). But cirques outside areas of ice sheet glaciation could develop at glacial maxima, whereas those within probably developed largely as ice built up.

References

Bennett, M.R., Huddart, D. and Glasser, N.F. (1999) Large-scale bedrock displacement by cirque glaciers, *Arctic, Antarctic and Alpine Research* 31(1), 99–107.

Derbyshire, E. and Evans, I.S. (1976) The climatic factor in cirque variation, in E. Derbyshire (ed.) *Geomorphology and Climate*, 447–494, Chichester: Wiley.

Evans, I.S. (1997) Process and form in the erosion of glaciated mountains, in D.R. Stoddart (ed.) *Process and Form in Geomorphology*, 145–174, London: Routledge.

——(1999) Was the cirque glaciation of Wales time-transgressive or not? *Annals of Glaciology* 28, 33–39.

Evans, I.S. and Cox, N.J. (1995) The Form of Glacial Cirques in the English Lake District, Cumbria, *Zeitschrift für Geomorphologie* 39(2), 175–202.

Gardner, J.S. (1987) Evidence for headwall weathering zones, Boundary Glacier, Canadian Rocky Mountains, *Journal of Glaciology* 33(113), 60–67.

Gordon, J.E. (1977) Morphometry of cirques in the Kintail–Affric–Cannoch area of N.W. Scotland, *Geografiska Annaler* A59, 177–194.

——(2001) The corries of the Cairngorm Mountains, *Scottish Geographical Journal* 117(1), 49–62.

Graf, W.L. (1976) Cirques as glacier locations, *Arctic and Alpine Research* 8(1), 79–90.

Haynes, V.M. (1968) The influence of glacial erosion and rock structure on corries in Scotland, *Geografiska Annaler* A50, 221–234.

Hooke, R. Le B. (1991) Positive feedbacks associated with erosion of glacial cirques and overdeepenings, *Geological Society of America Bulletin* 103, 1,104–1,108.

Iverson, N.R. (1991) Potential effects of subglacial water pressure on quarrying, *Journal of Glaciology* 37, 27–36.

Lewis, W.V. (ed.) (1960) Norwegian cirque glaciers, *Royal Geographical Society Research Series* No. 4. London.

Olyphant, G.A. (1981) Allometry and cirque evolution, *Geological Society of America Bulletin* 92, Part I, 679–685.

Richardson, C. and Holmlund, P. (1996) Glacial cirque formation in Northern Scandinavia, *Annals of Glaciology* 22, 102–106.

SEE ALSO: aspect and geomorphology; freeze–thaw cycle; glacial erosion; glacial protectionism; glacier

IAN S. EVANS

CLAY-WITH-FLINT

The chalklands of southern Britain (and northern France) are mantled over extensive areas by a group of deposits called clay-with-flint (*argile à silex*) (Laignel *et al.* 2002).

They are highly variable in composition, ranging 'from heavy reddish brown clays with large unworn flint nodules to almost stoneless yellow or white sands, yellowish to reddish brown silt loams, brightly mottled (red, lilac, green and white) stoneless clays, and beds of rounded flint pebbles' (Catt 1986: 151). Early English geologists tended to regard them as an insoluble residue, left after a long period of dissolution and weathering of the chalk. However, although some of the constituent material of clay-with-flint may have been derived from this source, it is not an adequate explanation of the variability of the material nor of the presence of miscellaneous types of clay, sand and flint shape. Much of it is probably derived and reworked from Palaeogene beds and other Cenozoic deposits, as Jukes-Browne (1906) so astutely recognized.

References

Catt, J.A. (1986) The nature, origin and geomorphological significance of clay-with-flints, in G. de G. Sieveking and M.B. Hart (eds) *The Scientific Study of Flint and Chert*, 151–159, Cambridge: Cambridge University Press.

Jukes-Brown, A.J. (1906) The clay-with-flints: its origin and distribution, *Quarterly Journal of the Geological Society of London* 62, 132–164.

Laignel, B., Quesnel, F. and Meyer, R. (2002) Classification and origin of the clay-with-flints of the western Paris Basin (France), *Zeitschrift für Geomorphologie* 46, 69–91.

A.S. GOUDIE

CLIFF, COASTAL

A cliff is a steep slope (usually >40°, often vertical and sometimes overhanging), exposing rock formations (Plate 24). Most coastal cliffs have been produced by wave ABRASION at the cliff base, but some have been formed by faulting or earlier fluvial or glacial erosion.

Cliffs cut in unconsolidated formations are known as Earth cliffs (May 1972), and those at

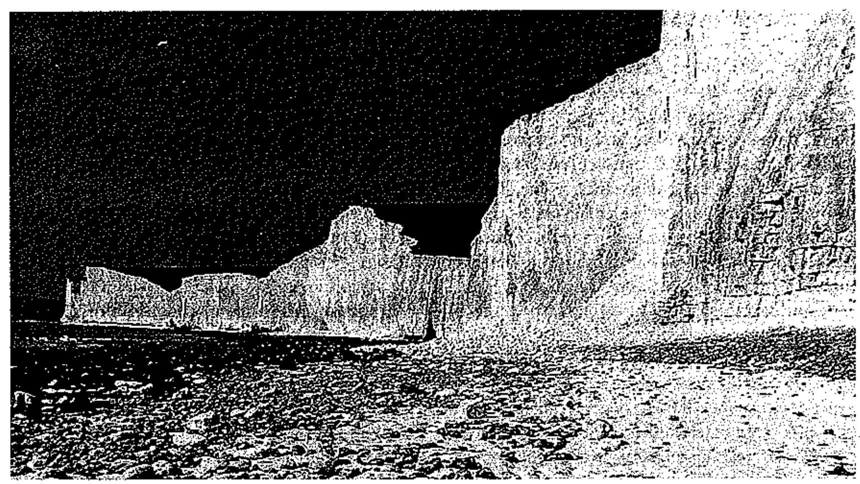

Plate 24 Chalk cliffs at Seven Sisters, Sussex, England, retreat as the result of wave abrasion, but are also influenced by solution, bioerosion and rock falls due to freeze–thaw effects and groundwater discharge

the seaward ends of glaciers ice cliffs. Hard rock cliffs, which change very slowly, have been relatively neglected in coastal research. In humid regions soil and vegetation may cover coastal slopes, except on actively receding cliff faces. Vegetated bluffs are not necessarily stable: on the Oregon coast they are cut back as cliffs during occasional severe storms or tsunamis, and then revegetate.

Cliffs rising 100–500 m above sea level are termed high cliffs, and those >500 m (as in Peru and western Ireland) megacliffs (Guilcher 1966). Cliffs less than a metre high are termed microcliffs.

Coastal cliffs recede as the result of basal marine erosion accompanied by subaerial erosion of the cliff face. A sharp angle or notch (see NOTCH, COASTAL) at the cliff base generally indicates active marine erosion. Some cliff profiles are of uniform gradient, others concave or convex, or a combination of these. Concave profiles occur where subaerial erosion exceeds marine erosion and convex profiles where marine erosion has been dominant (Emery and Kuhn 1982), but cliff profiles are also related to the position and inclination of resistant strata. A resistant caprock forms bold cliffs, hard outcrops in the cliff face produce ledges, and a resistant formation at the cliff base slows marine erosion (Figure 25A). A seaward dip facilitates landslides, horizontal strata may form stepped profiles and a landward dip produces an escarpment cliff (Figure 25B). Slope-over-wall profiles may be related to weak above resistant formations or an undercut seaward dipslope (Figure 25C: 1, 2). Joints, bedding planes, faults and intrusions influence cliff morphology, and lateral changes in lithology result in changes in cliff profiles, as on Triassic sandstones and clays in south-east Devon, England. On limestone coasts marine erosion exposes caves and cauldrons produced by earlier karstic dissection.

Cliff outlines in plan are also related to geological structure, with headlands where resistant formations outcrop at the cliff base and bays where weaker formations are excavated by marine erosion; headlands often coincide with ridges and bays with valleys. The Dorset coast, east of Weymouth in southern England illustrates these relationships (Bird 1995).

Cliff-base erosion is achieved by wave quarrying, which dislodges and removes rock material, and abrasion where waves throw sand or gravel against the cliff base. Cliff outcrops may disintegrate as the result of WETTING AND DRYING WEATHERING of surfaces subject to spray, splash and rainwash, or SALT WEATHERING where salt crystallizes from sea splash, notably on arid coasts. Solution by runoff, seepage, spray and sea water contributes to cliff-base erosion, particularly on limestone coasts where distinctive flat-floored solution notches may form, in contrast with sloping ramps where wave abrasion is dominant. Bioerosion (by plants and animals that live on the cliff and shore) also contributes.

Cliff faces may be indurated by calcareous or ferruginous compounds precipitated from groundwater seepage, forming crusts that eventually crack and exfoliate, exposing uncemented rock. Cliff faces are also indurated by carbonates precipitated from sea splash, particularly on headlands. Downwashed sediment may adhere to a cliff face as stalactitic structures, notably on limestone and AEOLIANITE (Hills 1971). By contrast, fine-grained sediment winnowed from a cliff face by onshore winds has been deposited as a cliff-top levee on the Port Campbell coast in south-eastern Australia (Baker 1943).

As a cliff is undercut it may collapse, producing a debris fan below a fresh rock scar. Sediment yield from cliffs depends on the rate of recession and the effects of weathering and erosion. Accumulation of sediment at the cliff base slows recession, but usually the debris fan is dispersed by erosion, and when it has been removed basal undercutting resumes.

MASS MOVEMENTS occur on cliffs where the groundwater load becomes excessive, where stresses develop as the result of freeze and thaw, where a massive caprock exerts pressure on underlying weaker formations, or where there is expansion or base exchange, weakening clay minerals. Breakaways develop at the cliff crest where masses of rock topple down the cliff, and slumping produces irregular topography as rock outcrops disintegrate and material slides, flows or creeps down the slope towards a basal receding cliff. Such cascading systems, with instability transmitted upward to the cliff crest, occur on the Dorset coast (Brunsden and Jones 1980). In Oregon coastal landslides in weathered rock are commoner in winter, when stronger wave action attacks formations saturated by heavy rain, but may also be triggered by tectonic movements or tsunamis generated along the nearby plate edge.

Some cliffs descend to SHORE PLATFORMS cut by marine erosion and weathering processes as the cliff recedes; others are fronted by irregular rocky shores, particularly where the geological formations are of intricate structure with resistant elements; others (plunging cliffs) continue below sea level, either because of partial marine submergence (where they descend to submerged coastlines) or because they formed by faulting, glaciation or vulcanicity.

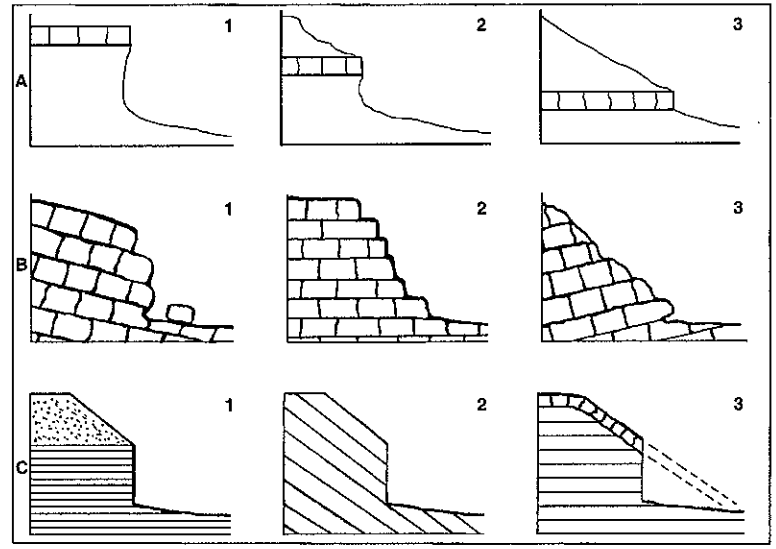

Figure 25 A, the effects of a resistant formation on cliff profiles; B, variations related to the dip of strata; C, slope-over wall cliffs: 1, related to lithology; 2, related to structure; 3, retaining a slope formed by periglacial solifluction

Some cliffs are actively receding; others are inactive behind persisting basal talus or a prograding beach, or because of lowering of sea level. Inactive cliffs may decline into subaerially shaped slopes which become vegetated. Cliffs stranded by land uplift or sea-level lowering become bluffs behind emerged beaches and shore platforms.

Rates of cliff recession vary with cliff height, rock resistance, structure, weathering and exposure to wave attack. They are usually reported as annual averages, but are generally episodic, related to occasional storms or mass movements. Rapid cliff recession ($>1\,\mathrm{m\,yr^{-1}}$) occurs on soft rock formations, and rates of $>100\,\mathrm{m\,yr^{-1}}$ have been reported on cliffs in volcanic ash and arctic tundra deposits (humates with melting ice), but some hard rock cliffs have shown little or no recession in the period (up to 6,000 years) that the sea has stood at its present level.

Where cliff recession has been slow, features inherited from earlier environments may persist. Examples of this are the slope-over-wall profiles on the Atlantic coasts of Britain, where the slope (which may be convex, a straight bevel or concave) is mantled by earthy gravel (termed Head) formed by periglacial SOLIFLUCTION in cold phases of the Pleistocene, and the wall is a receding undercliff (Figure 25C: 3): the proportion of slope to wall diminishes as exposure to wave attack increases. Active periglaciation forms steep slopes of angular debris on arctic coasts, as on Baffin Island in Canada. In northern Britain and Scandinavia the periglacial slope gives place to slopes formed by glacial erosion or deposition, also undercut by Holocene marine erosion. In the humid tropics slope-over-wall profiles occur where a coastal slope on deeply weathered rock has been undercut by marine erosion, and in arid regions the undercut coastal slope may have been a pediment.

Cliff recession is likely to accelerate (and coastal landslides become more frequent) during a rising relative sea level, and when storminess increases in coastal waters: protective beaches diminish, and wave attack on the cliff base becomes stronger and more sustained.

Human impacts on cliffs include stabilization by the building of basal sea walls or boulder ramparts to halt coastline retreat, the grading, vegetating or concreting of cliff faces, and the introduction of drains to hasten groundwater discharge. By contrast, cliffs become more unstable as the result of the reduction of beaches (when beach sand or gravel are extracted), an increase in groundwater load and levels (when previously dry cliff-top terrain is irrigated) and cliff-top loading by buildings and other structures.

References

Baker, G. (1943) Features of a Victorian limestone coastline, *Journal of Geology* 51, 359–386.
Bird, E.C.F. (1995) *Geology and Scenery of Dorset*, Bradford-on-Avon: Ex-Libris Press.
Brunsden, D. and Jones, D.K.C. (1980) Relative timescales in formative events in coastal landslide systems, *Zeitschrift für Geomorphologie, Supplementband* 34, 1–19.
Emery, K.O. and Kuhn, G.G. (1982) Sea cliffs: their processes, profiles and classification, *Geological Society of America Bulletin* 93, 644–654.
Guilcher, A. (1966) Les grandes falaises et mégafalaises des côtes sud-ouest et ouest de l'Irlande, *Annales de Géographie* 75, 26–38.
Hills, E.S. (1971) A study of cliffy coastal profiles based on examples in Victoria, Australia, *Zeitschrift für Geomorphologie* 15, 137–180.
May, V.J. (1972) Earth cliffs, in R.S.K. Barnes (ed.) *The Coastline*, 215–235, New York: Wiley.

Further reading

Sunamura, T. (1992) *Geomorphology of Rocky Coasts*, Chichester: Wiley.
Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Clarendon.

SEE ALSO: slope, evolution

ERIC C.F. BIRD

CLIMATIC GEOMORPHOLOGY

The part of the discipline that seeks to explain the form and distribution of landforms in terms of climate. It developed during the period of European colonial expansion and exploration at the end of the nineteenth century, when unusual and often spectacular landforms were encountered in deserts, polar regions and the humid tropics. In addition, it was a time when regionalization and classification were major endeavours in geography and cognate subjects. Attempts at climatic, soil and vegetation classifications were being made by scientists like Köppen, Dokuchayev and Schimper. They sought to understand the regional patterns of the phenomena they were classifying, and climate was seen as a major control at their scale of investigation.

In the USA, W.M. Davis recognized 'accidents', whereby non-temperate and non-humid climatic regions were seen as deviants from his normal cycle of erosion and he introduced, for example, his arid cycle (Davis 1905). Some (see Derbyshire 1973) regard Davis as one of the founders of climatic geomorphology, although the leading French climatic geomorphologists Tricart and Cailleux (1972) criticized Davis for his neglect of the climatic factor in landform development. Much important work was undertaken on dividing the world into climatic zones (morphoclimatic regions) with distinctive landform assemblages, in France (e.g. Birot 1968), in Germany (e.g. Büdel 1982) and in New Zealand (Cotton 1942). This version of geomorphology was seen as essentially geographical (Holzner and Weaver 1965).

In the later years of the twentieth century the popularity of climatic geomorphology became less as certain limitations became apparent (see Stoddart 1969).

(1) Much climatic geomorphology was based on inadequate knowledge of rates of processes and on inadequate measurement of process and form. Assumptions were made that, for example, rates of chemical weathering were high in the humid tropics and low in cold regions, whereas subsequent empirical studies have shown that this is far from inevitable.
(2) Some of the climatic parameters used for morphoclimatic regionalization were meaningless or crude from a process viewpoint (e.g. mean annual air temperature).
(3) Macroscale regionalization was seen as having little inherent merit and ceased to be a major goal of geographers, who eschewed 'placing lines that do not exist around areas that do not matter'.
(4) Conversely, and paradoxically, climatic geomorphology had a tendency to concentrate on bizarre forms found in some 'extreme' environments rather than on the overall features of such areas.
(5) Many landforms that were supposedly diagnostic of climate (e.g. pediments in arid regions or inselbergs in the tropics) are either very ancient relict features that are the product of a range of past climates or they have a form that gives an ambiguous guide to origin.
(6) The impact of the large, frequent and rapid climatic changes of the Quaternary and of the very different climates of the Tertiary has disguised any simple climate–landform relationship. For this reason, Büdel (1982) attempted to explain landforms in terms of fossil as well as present-day climatic influences (Table 8). He recognized that landscape was composed of various 'relief generations' and saw the task of what he termed 'climato-genetic geomorphology' as being to recognize, order and distinguish these relief generations, so as to analyse today's highly complex relief.

Table 8 Büdel's morphogenetic zones of the world

Zone	Present climate	Past climate	Active processes (fossil ones in brackets)	Landforms
(1) Of glaciers	Glacial	Glacial	Glaciation	Glacial
(2) Of pronounced valley formation	Polar, tundra	Glacial, polar, tundra	Frost, mechanical weathering, stream erosion (glaciation)	Box valleys, patterned ground, etc.
(3) Of extra-tropical valley formation	Continental, cool temperate	Polar, tundra continental	Stream erosion (frost processes, glaciation)	Valley
(4) Of subtropical pediment and valleys formation	Subtropical (warm; wet or dry)	Continental, subtropical	Pediment formation (stream erosion)	Planation surfaces and valleys
(5) Of tropical plantation surface formation	Tropical (hot; wet or wet–dry)	Subtropical, tropical	Planation, chemical weathering	Planation surfaces and laterites

Although these tendencies have tended to reduce the relative importance of traditional climatic geomorphology, notable studies still appear that look at the nature of landforms and processes in different climatic settings (e.g. M. Thomas 1994 on the humid tropics; D. Thomas 1998 on arid lands; and French 1999 on periglacial regions). In addition, a concern with GLOBAL WARMING and its geomorphological impact leads to a renewed concern with climate-landform links.

References

Birot, P. (1968) *The Cycle of Erosion in Different Climates*, London: Batsford.
Büdel, J. (1982) *Climatic Geomorphology*, Princeton, NJ: Princeton University Press.
Cotton, C.A. (1942) *Climatic Accidents in Landscape-making*, Christchurch: Whitcombe and Tombs.
Davis, W.M. (1905) The geological cycle in an arid climate, *Journal of Geology* 13, 381–407.
Derbyshire, E. (ed.) (1973) *Climatic Geomorphology*, London: Macmillan.
French, H.M. (1999) *The Periglacial Environment*, 2nd edition, London: Longman.
Holzner, L. and Weaver, G.D. (1965) Geographical evaluation of climatic and climato-genetic geomorphology, *Annals of the Association of American Geographers* 55, 592–602.
Stoddart, D.R. (1969) Climatic geomorphology, in R.J. Chorley (ed.) *Water, Earth and Man*, 473–485, London: Methuen.
Thomas, D.S.G. (ed.) (1998) *Arid Zone Geomorphology*, 2nd edition, Chichester: Wiley.
Thomas, M.F. (1994) *Geomorphology in the Tropics: A Study of Weathering and Denudation in Low Latitudes*, Chichester: Wiley.

Tricart, J. and Cailleux, A. (1972) *Introduction of Climatic Geomorphology*, London: Longman.

Further reading

Elorza, M.G. (2001) *Geomorfología climática*, Barcelona: Ediciones Omega.

A.S. GOUDIE

CLIMATO-GENETIC GEOMORPHOLOGY

Climato-genetic geomorphology is the systematic field investigation of landforms in a certain area according to their evolution. Many different methods may be applied, but the basis for it is the observation of an assemblage of rested relief elements. There are two roots to climato-genetic geomorphology. First, the more palaeoforms were acknowledged, it became clear that their systematic investigation was necessary, not only to explain the relief but also to estimate their influence on recent processes. Second, the system of 'klimatische Geomorphologie' was developed. It is the basis for relief forming processes or process fabric, which is applied to the different RELIEF GENERATIONS, the constituents of climato-genetic geomorphology.

The terminology is not very clear for originally the term 'klimatische Geomorphologie' was introduced to differentiate it from tectonic or structural geomorphology. However, it was misleading. Dynamic geomorphology would have been a much better term, as it is the study of processes mainly at the medium scale. 'Climatic geomorphology' is the literal translation, but this has the very different aim of relating landforms to climatic or hydrological data. 'Klimatische Geomorphologie' investigates the relief forming processes in a certain MORPHOGENETIC REGION. Climatic geomorphology looks more for single forms or processes. More or less similar to 'klimatische Geomorphologie' are 'research in a morphoclimatic zone', 'climato-geomorphology', 'forms of morphoclimates' or 'DYNAMIC GEOMORPHOLOGY'. For the evolution of landforms there are the words 'klimatische Morphogenese' (literally climatic morphogenesis) and 'klimagenetische Geomorphologie' (literally climato-genetic morphology). Thus the position of the words geomorphology, climate and genetic might change without any generally agreed special connotations.

There is a distinction though between processes and evolution in the terms. It seems that there is a difference in the English and German use of the word 'genetic' in geomorphology, that is, development and historical outline of natural phenomena.

Processes were deduced rather early on in the search for an explanation of landforms, and their relation to exogene (i.e. climate controlled force) was acknowledged. In Europe the work of glaciers was studied on recent examples and similar landforms and deposits were classified accordingly (ACTUALISM). In the west of the USA early research detected the specific processes of the arid zone. Palaeoforms have been increasingly acknowledged since the early twentieth century. The systematic approach to climatic geomorphology dates from 1948. After the Second World War the overseas research of palaeoclimatology (see PALAEOCLIMATE), deduced from morphological features like moraines and solifluction forms, increased. All these research efforts were the basis for the concept of the development of RELIEF GENERATIONS. There are several possibilities for applying this concept besides the explanation of relief evolution. It may serve to control erosion rates, especially their extrapolation and the distinction of human accelerated rates. On the other hand, the extension and preservation of the different relief generations shows the intensity and specific location of recent land forming activity. This is a good basis for applied questions like soil erosion. In connection with ecological studies, relief generations are a basis for the spatial extent of investigated features, e.g. the distribution of soil types.

The recent process fabric is either observed directly or deduced from fresh landform scars after catastrophic events. Similar forms of different size, and sequences, are extrapolated to get an idea about intensity, recurrence and the forming power of special processes and their interrelation. The relation between denudation and linear erosion is investigated as well as between erosion and deposition. This is counterchecked by the known facts of climatic change and tectonic movements, which give an estimate of the change of the processes. As the process fabric is a systematic combination of single geomorphological activities one can ask for completeness of processes as well as forms. A simple example may illustrate this: in the northern foreland of the Alps the rivers now carry sand and show a rather low activity. The slopes are more or less undisturbed. Thus the younger process fabric is not strong and not widespread. There is a large amount which remains unexplained, which from the analysis of the forms is easily classified as moraines. If a soil is developed on them, they are inactive. Moraines are known from the surroundings of recent glaciers in their form and sedimentary structure. Connected landforms are outwash plains in front of them and overdeepening to the rear of the moraines. With this form assemblage the older process fabric can be extended beyond the moraines. Gravel terraces in front of them are of fluvioglacial origin, while lakes to the rear are a sign of glacial scour. There is a feedback in the analysis of forms and processes. Many more details and several stages of the advancing and retreating glaciers have been classified and mapped, e.g. for the Inn Chiemsee glacier by Carl Troll.

With the advancement of knowledge about relief generations it became clear that older forms are widely distributed. There are a few landscapes, like young volcanoes or badlands, which consist only of one relief generation.

Therefore fieldwork should start with the oldest forms and look for the nested younger form assemblage. By interpolation, the younger process fabric is derived. There are the same feedback mechanisms as those named above. This method has two advantages: the existence of remaining unexplained phenomena can be avoided, which one is always inclined to keep as low as possible. Second 'Mehrzeitformen' (i.e. forms shaped in different climates) are more easily detected. It is self-evident that there is a slight change to all the forms of the older relief generation (e.g. the removal of the soil cover of an old plain), but there are a few forms which were noticeably changed by younger processes, like blockfields in the mid-latitudes (cf. RELIEF GENERATIONS).

Climato-genetic geomorphology works not only by analysing landforms, but also uses a wide range of other methods. As a genetic science the connection to the well-developed soil science is especially close. For the tropical zone soil analysis in the field and in the laboratory can solve problems of allochthonous or autochthonous weathering, of relative age, and especially of palaeofeatures. In the humid mid-latitudes, relics of tropical weathering, the periglacial cover of solifluction and loess are counterchecks for the distinction of relief generations. All direct and indirect dating methods are helpful.

Further reading

Bremer, H. (2002) Tropical weathering, landforms and geomorphological processes: fieldwork and laboratory analysis, *Zeitschrift für Geomorphologie* 46, 273–291.
Büdel, J. (1977) *Klima-Geomorphologie*, Borntraeger: Berlin. Translated by L. Fischer and D. Busche (1982) *Climatic Geomorphology*, Princeton: Princeton University Press.

HANNA BREMER

COASTAL CLASSIFICATION

Coastal classification is the grouping of similar coastal features in categories that distinguish them from dissimilar features. The aim is to elucidate the relationships between coastal landforms and processes and to understand coastal evolution. Simple classifications are implicit in the topics identified in chapter headings in coastal textbooks, and when coastal features are categorized and shown on maps of coastal morphology.

Some attempts to classify coastal landforms (including shores and shoreline features) have been genetic, based on the origin of the landforms, rather than descriptive (e.g. cliffed coasts, delta coasts, mangrove coasts). The difficulty is that genetic classifications can only be applied when the mode of origin of coastal landforms is known, and as only a small proportion of the world's coastline has been investigated in sufficient detail to determine evolution such classifications remain somewhat speculative. Certainly the assumption that particular types or associations of landforms can be used as indicators of particular modes of origin can be misleading, for some coastal landforms (e.g. barrier islands, beach ridges, cuspate forelands, shore platforms) may evolve in more than one way: a phenomenon termed multicausality (Schwartz 1971). Various kinds of coastal classification are now described, with references.

Atlantic and Pacific type coasts

Suess (1906) distinguished Atlantic coasts, which run across the general trend of geological structures, from Pacific coasts, which run parallel to structural trends. The former are characteristic of the Atlantic shores of Britain and Europe; the latter of the Pacific coasts of North and South America.

Cliffed coastlines that transgress geological structures are termed discordant, whereas those that follow the strike of a particular geological formation are termed concordant.

Classification and plate tectonics

Inman and Nordstrom (1971) devised a geophysical classification based on PLATE TECTONICS, recognizing that the Earth's crust is a pattern of plates separated by zones of spreading and zones of convergence, with plate margins moving at rates of up to $15\,\mathrm{cm\,yr^{-1}}$. They contrasted subduction coasts, where one plate is passing beneath another, with trailing-edge coasts on a diverging plate margin and marginal sea coasts on the lee side of island arcs, and described features characteristic of each of these. It was a broad-scale classification, dealing with first-order (continental) features (*c.*1,000 km long × 100 km wide × 10 km high).

Coasts of submergence and emergence

Gulliver (1899) distinguished coasts formed by submergence from coasts formed by emergence. This was developed into a genetic classification by Johnson (1919), who described coastlines (he used the American term shorelines) of submergence, coastlines of emergence, neutral coastlines (with forms due neither to submergence nor emergence, but to deposition, e.g. delta coastlines, alluvial plain coastlines, glacial outwash coastlines and volcanic coastlines) and compound coastlines (with an origin combining two or more of the preceding categories). Most coasts fall into the compound category, because they show evidence of both emergence, following high sea levels in interglacial phases of the Pleistocene, and submergence, due to the Late Quaternary (Flandrian) marine transgression.

Classification based on climate

Aufrère (1936) proposed a coastal classification based on climate, which distinguished coasts with a permanent ice cover (no marine processes), coasts with a seasonal ice cover (seasonal marine processes and abundant sediment from glacial sources), temperate humid coasts (as in Europe), tropical humid coasts (with abundant fluvial sediment in deltas and coastal plains), arid coasts (without rivers; marine sediments dominant) and semi-arid coasts (some river features; SABKHAs). The global distribution of coastal climates shows sector variations related to latitude and wind regime with coastwise transitions that are generally gradual, although there are rapid transitions from humid tropical to arid within comparatively short distances in Ecuador and Colombia, in west Africa and northern Madagascar.

Classification based on coastal processes

Variations in coastal processes effective around the world's coastline were discussed by Davies (1980), who defined and mapped swell and storm wave environments, coasts subject to trade winds, monsoons and tropical cyclones, the distribution of high, moderate and low wave energy coasts, tidal types (semi-diurnal, mixed and diurnal) and mean maximum tide ranges divided into microtidal (<2 m), mesotidal (2–4 m) and macrotidal (>4 m), to which may be added megatidal (>6 m).

Initial and subsequent coasts

A distinction can be made between initial forms, which existed when the present relative levels of land and sea were established and marine processes began work (on most coasts about 6,000 years ago) and sequential forms, those that have since developed as the result of marine action. Shepard (1976) devised a classification on this basis, making a distinction between primary coasts shaped largely by non-marine agencies and secondary coasts that owe their present form to marine action. It was essentially a genetic classification, with descriptive detail inserted to clarify the subdivisions, and it recognized that, because of the worldwide Late Quaternary marine transgression, the sea has not long been at its present level relative to the land, so that many coasts have been little modified by marine processes.

Shepard's aim was to devise a classification that would prove useful in diagnosing the origin and history of coastlines from a study of charts and air photographs, but it is dangerous to assume that the origin and history of a coast can be deduced from such evidence without field investigation. A straight coast may be produced by deposition, faulting, emergence of a featureless seafloor or submergence of a coastal plain; an indented coast by submergence of an undulating or dissected land margin, emergence of an irregular seafloor, differential marine erosion of hard and soft outcrops along the coast or transverse tectonic deformation (folding and faulting) of the land margin. It is doubtful whether configuration can be taken as a reliable indicator of coastal evolution.

Leontyev *et al.* (1975) also considered initial and sequential forms (using the cycle of youth, maturity and old age) in a classification based on coasts not changed by the sea, coasts formed by abrasion or accumulation, and a combination of the two.

Stable and mobile coasts

Cotton (1952) made a distinction between coasts of stable and mobile regions, stable regions being those that escaped the Quaternary tectonic movements that have affected mobile regions, especially around the Pacific rim, where they still continue. On the coasts of stable regions he separated those dominated by features produced by Late Quaternary marine submergence from those dominated by inherited (mainly Pleistocene) features preserved by earlier emergence. On the coasts of mobile regions he separated those where the effects of Late Quaternary marine submergence have not been counteracted by recent uplift of the land from those where recent uplift of the land has caused emergence.

Morphological classification

De Martonne (1909) used a morphological distinction between steep and flat coasts as a basis for classification, suggesting a number of subtypes, some descriptive (estuary coasts, skerry coasts), others genetic (fault coasts, glacially sculptured coasts). Ottmann (1965) followed a similar approach, recognizing three categories of cliffed coast (cliffs plunging to oceanic depths, cliffs with shore platforms and cliffs plunging to submerged platforms), partially submerged uncliffed coasts, and low depositional coasts behind gently shelving seafloors.

Zenkovich (1967) classified depositional coastal features into five categories: attached forms (including beaches and cuspate forelands), free forms (including spits), barriers, looped forms (including tombolos) and detached forms (including barrier islands).

Geology in coastal classification

Russell (1967) advocated classification of rocky coasts on the basis of geology and structure, noting the striking similarity of features developed on crystalline rocks, irrespective of their climatic and ecological environments: granites that outcrop on parts of the coasts of Scandinavia, southwest Australia, South Africa and Brazil all show similar domed surfaces related to large-scale spalling and conspicuous joint-control. Limestones (including chalk and coral), basalts and sandstones also show distinctive kinds of coastal landforms. Bedrock coasts are commoner in cold, arid and temperate regions than in the humid tropics, where there has been deep weathering and depositional aprons are extensive.

Advancing and receding coasts

A coastline may advance because of coastal emergence and/or progradation by deposition, or retreat because of coastal submergence and/or retrogradation by erosion. Valentin (1952) used this analysis as the basis for a system of coastal classification that could be shown on a world map. Coasts that had advanced were divided into those produced by emergence, by organic deposition (mangroves, coral) and by inorganic deposition (marine and fluvial), while coasts that had retreated were divided into those produced by submergence of glaciated landforms and fluvially eroded landforms and those shaped by marine erosion. Bloom (1965) elaborated Valentin's scheme by considering historical evolution where the response to emergence, submergence, erosion and deposition has varied through time. Thus on the Connecticut coast, where radiocarbon dates from buried peat horizons have yielded a chronology of relative changes of land and sea level in Holocene times, there is evidence that at some stages the sea gained on the land during submergence, even though deposition continued, while at other stages deposition was sufficiently rapid to prograde the land during continuing submergence: at present there is widespread erosion on the seaward margins of saltmarshes, possibly because of resumed submergence.

The advantage of such non-cyclic classifications is that they pose problems and stimulate further research instead of trying to fit observed features into presupposed evolutionary sequences.

Composite classifications

McGill (1958) produced a map of the world's coastline which showed the major landforms of the coastal fringe, 8–16 km wide. This was a composite classification in which major coastal landforms were classified in terms of lowland or upland hinterlands, with additional information on selected features (constructional or destructional) in the backshore, foreshore and offshore zones, categorized by the agent responsible: sea, wind, coral or vegetation.

Artificial coastlines

Little attention has been given in coastal classifications to the fact that long sectors of coastline have become artificial during recent decades, partly as the result of engineering works designed to combat erosion and partly as a consequence of embanking or infilling to extend coastal land. On developed coasts the proliferation and extension of anti-erosion works, notably sea walls and boulder ramparts, has resulted in large proportions of artificial coastline: 85 per cent in Belgium, 51 per cent in Japan, 38 per cent in England. Coastal land has been artificially extended on a large scale in Singapore, Hong Kong, Tokyo Bay in Japan, western Malaysia and the Netherlands. The category of artificial coastlines is increasing rapidly, and much more of the world's coastline will become artificial as attempts are made to halt submergence and erosion.

References

Aufrère, L. (1936) Le rôle du climat dans l'activité morphologique littorale, *Proceedings, 14th International Geographical Congress*, Warsaw, 2, 189–195.
Bloom, A.L. (1965) The explanatory description of coasts, *Zeitschrift für Geomorphologie*, 9, 422–436.
Cotton, C.A. (1952) Criteria for the classification of coasts, *Proceedings 17th Conference, International Geographical Union*, Washington, 315–319.
Davies, J.L. (1980) *Geographical Variation in Coastal Development*, London: Longman.
De Martonne, E. (1909) *Traité de Géographie Physique*, Paris: Colin.
Gulliver, F.P. (1899) Shoreline topography, *Proceedings, American Academy of Arts and Sciences*, 34, 151–258.
Inman, D.L. and Nordstrom, C.E. (1971) On the tectonic and morphologic classification of coasts, *Journal of Geology* 79, 1–21.
Johnson, D.W. (1919) *Shore Processes and Shoreline Development*, New York: Wiley.
Leontyev, O.K., Nikiforov, L.G. and Safyanov, G.A. (1975) *The Geomorphology of the Sea Coasts* (in Russian), Moscow: Moscow University.
McGill, J.T. (1958) Map of coastal landforms of the world, *Geographical Review* 48, 402–405.
Ottmann, F. (1965) *Introduction a la Géologie Marine et Littorale*, Paris: Masson.
Russell, R.J. (1967) *River Plains and Sea Coasts*, Berkeley, CA: University of California Press.
Schwartz, M.L. (1971) The multiple causality of barrier islands, *Journal of Geology* 79, 91–93.
Shepard, F.P. (1976) Coastal classification and changing coastlines, *Geoscience and Man* 14, 53–64.
Suess, E. (1906) *The Face of the Earth*, Oxford: Clarendon Press.
Valentin, H. (1952) *Die Küsten der Erde*, Petermanns Geographische Mitteilungen, 246.
Zenkovich, V.P. (1967) *Processes of Coastal Development*, Edinburgh: Oliver and Boyd.

Further reading

Bird, E.C.F. (2000) *Coastal Geomorphology: An Introduction*, Chichester: Wiley.
Schwartz, M.L. (ed.) (1982) *The Encyclopedia of Beaches and Coastal Environments*, Stroudsburg, PA: Hutchinson Ross.

SEE ALSO: coastal geomorphology; global geomorphology

ERIC C.F. BIRD

COASTAL GEOMORPHOLOGY

The industrial, recreational, agricultural and transportational activities of growing human populations are exerting enormous pressures on coastal resources. To manage these activities in the least detrimental way, we need to have a better understanding of the dynamic nature of coastal landforms and the operation and interaction of marine and terrestrial processes. Differences in climate, changes in relative SEA LEVEL, wave environments, tides, winds, the morphology, structure and lithology of the hinterland, terrestrial and marine sediment sources, human activity and numerous other factors provide almost infinite variety to coastal scenery around the world. Coastal regions consist of a mosaic of diverse elements, some of which are contemporary, whereas others are ancient vestiges of periods when climate and sea level may have been similar or different from today's. Small-scale elements of depositional coasts, which can experience rapid changes in morphology, may attain a rough state of balance with their environmental conditions, but other features – particularly on hard rock coasts – require long periods to adjust to changing conditions. Furthermore, even if environmental conditions remain constant, individual coastal landforms still have to adjust to slow changes in the morphology of the coast itself. For example, whereas the profiles of sandy BEACHes respond fairly quickly to changing wave conditions, they may also have to adjust slowly to long-term changes in coastal configuration, sediment budgets, offshore gradients, climate, sea level and increasingly the effects of human interference.

Coastal classification

There have been many attempts to classify coasts, although none are entirely satisfactory. Most COASTAL CLASSIFICATIONS use at least two of three basic variables: the shape of the coast; changes in relative sea level; and the effect of marine processes. Some classifications are genetic, others are descriptive and others combine the two approaches. Genetic classifications are hindered by a lack of relevant data, however, and descriptive classifications, which have to accommodate an enormous variety of coastal types, tend to be cumbersome. Two classifications, which consider the nature of coastal environments and the effect of PLATE TECTONICS on coastal development, are particularly useful.

Davies (1972) proposed that coastal processes are strongly influenced by morphogenic factors that vary in a fairly systematic way around the world. Davies's morphogenic classification was based upon four major wave climates, although differences in coastal characteristics also reflect variations in tidal range, climate and many other factors. The highest WAVEs are usually generated in the storm belts of temperate latitudes. Beaches in storm wave environments tend to have dissipative or gently sloping and barred profiles, and the major constructional features are often composed of coarse clastic material. Constructional features are oriented more by local fetch than by the variable direction of the deep water waves, and mechanical wave erosion is important in the formation of cliffs (see CLIFF, COASTAL) and SHORE PLATFORMS. Long, low constructional waves dominate swell environments between the northern and southern storm wave belts. The beaches have berms, and they tend to be towards the steeper, reflective, non-barred end of the spectrum. The direction of longshore currents is more constant than in storm wave environments, and large, sandy constructional features are oriented toward the approaching swell. Mechanical wave erosion of cliffs and platforms is probably slower than in storm wave environments, and this, combined with warmer climates, makes CHEMICAL WEATHERING and biological WEATHERING more important in swell wave environments. Sheltered, enclosed seas and ice-infested waters are low energy environments. Waves are flat and constructional, and beaches have prominent berms. The orientation of sandy constructional features, which are common in partially enclosed seas, is largely determined by local fetch.

Plate tectonics provide a partial explanation for the distribution of a variety of coastal elements, although the degree of explanation decreases with the decreasing size of the feature. Inman and Nordstrom (1971) proposed that the morphology of the largest, or first-order, coastal elements can be attributed to their position on moving tectonic plates. Three main geotectonic classes were identified: continental and ISLAND ARC collision coasts form along the edges of converging plates; plate-imbedded or trailing edge coasts face spreading centres; and marginal sea coasts develop where island arcs separate and protect continental coasts from the open ocean. The structural grain of collision coasts is parallel to the shore and they are therefore fairly straight and regular. Tectonically mobile collision coasts have narrow continental shelves and high, steep hinterlands, often with flights of raised terraces. The high relief provides an abundant supply of sediment to the coast. Plate-imbedded or trailing edge coasts usually have hilly, plateau, or low hinterlands, and wide continental shelves. The structural grain may be at high angles to the coast, which can therefore be very indented. Marginal seacoasts range from low-lying to hilly, with wide to narrow shelves, and they are often modified by large rivers and RIVER DELTAS.

Coastal modelling

Models provide one of the best ways of investigating the poorly understood components of a coastal system. They provide insights into the interrelationships between and among variables, and they are indispensable in enhancing our efforts to monitor, manage, control and develop the coastal system and its associated resources.

Physical models are simplified and scaled representations of the real world. They can be used to control and isolate variables, to provide insights into phenomena not yet described or understood, to provide measurements to test theoretical results and to measure complicated phenomena that cannot be theoretically analysed. Coastal engineers have constructed a wide variety of fixed-bed hydraulic scale models to study the action of waves, tides and currents, and to assist in the design of coastal structures. Geologists and geomorphologists have used movable bed models to examine sediment transport and the dynamics and formation of bars (see BAR, COASTAL), barriers (see BARRIER AND BARRIER ISLAND) and beaches. Unlike natural oceanic waves, however, the shallow water waves generated in most wave tanks have no orbital kinetic energy and are nearly pure solitons. Physical models therefore have not been able to describe accurately the hydrodynamics and sedimentary processes operating in coastal systems, and the results obtained from them always have to be verified or corroborated with other evidence.

Because of their generality, versatility and flexibility, mathematical models are the most common type used by coastal workers. Unfortunately, however, our lack of knowledge of coastal processes and the frequent reliance on laboratory data to determine the value of coefficients, casts doubt on the applicability of many mathematical models to the real world. There are several types of mathematical model. Deterministic models, which are based on the principles of fluid mechanics, seem to work best in conjunction with laboratory experiments that allow parameters to be held constant while one is varied at a time. Simulation models involve the manipulation of process–response equations on computers, compressing years of coastal development in the prototype into minutes. This allows the behaviour of a system to be determined under a variety of situations and conditions, and to test the sensitivity of the system to changing input parameters. Statistical models can be used to study the relationships between a set of variables, and to verify possible relationships identified by theoretical models. To use equations derived from one area for predictive purposes in another, however, often requires the determination of a different set of coefficients.

Coastal inheritance

There is growing evidence that because interglacial sea levels were similar to today, contemporary coastal features often formed close to, or were superimposed on top of, their ancient counterparts. Although evidence of past sea levels and climates is generally easily obliterated in unconsolidated coastal deposits, many sandy coasts retain sedimentary and morphological elements of former environmental conditions. Coastal deposits from the last interglacial stage are being cannibalized in some areas to provide sediment for the construction and maintenance of modern coastal features, and barrier systems have sometimes developed on top of older Pleistocene barriers, or are located somewhat seaward of them. Most barrier islands on the German North Sea coast and in places on the Atlantic coast of the USA, for example, consist of a core of Pleistocene deposits, mantled by Holocene sediments. In south-eastern Australia, a distinct inner barrier of the last interglacial age is separated from an outer Holocene barrier by a lagoon and swamp tract. Pleistocene dunefields are adjacent, and probably under Holocene coastal dunes (see DUNE, COASTAL) in some places, especially in Australia and the Mediterranean, although they are generally absent in northern Europe, where most dunes were built at different stages during the Holocene. The presence of near-surface discontinuities shows that Holocene limestones, ranging from a few metres up to about 30 m in thickness, also form veneers over foundations of older reef-rock. The concept of INHERITANCE is particularly important on resistant rock coasts which have probably evolved very slowly during successive periods of high interglacial sea level. It has been demonstrated that some cliffs, sea caves, ramps (see RAMP, COASTAL) and shore platforms are at least last interglacial in age, and modelling suggests that many platforms have developed during interglacial stages during the middle and late Pleistocene.

Coastal management

Despite the problems associated with flooding, erosion, pollution and other hazards, and the increasing aesthetic and practical impetus for sustainable coastal management, rising populations and growing economic pressures are accelerating the pace of human interference and degradation on the world's coasts (Plate 25). We lack reliable models, however, that can be usefully employed by managers, planners and decision-makers for INTEGRATED COASTAL MANAGEMENT and to predict the effects of sea-level changes, human activities and other factors on the coast. The available field data on coastal changes are often of questionable reliability and usually too short-term to analyse the interaction of a large number of variables. Coastal changes are also frequently complex and non-linear (see NON-LINEAR DYNAMICS), and may reflect the interaction and exchange of sediment between the coast and the CONTINENTAL SHELF, and between the coast and the land, a relationship that is increasingly influenced by anthropological activities.

Long stretches of coastlines are now essentially artificial, with GROYNEs, breakwalls and other engineering structures (Plate 26). These structures

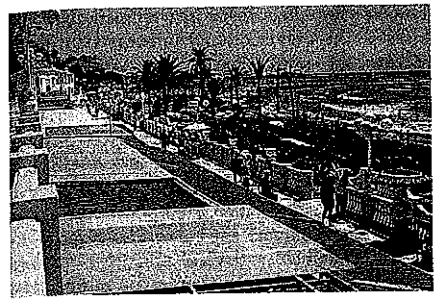

Plate 25 Crowded beach on the Costa del Sol, southern Spain

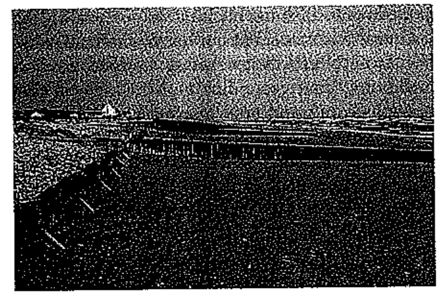

Plate 26 Groynes on Gold Beach, Normandy, France

are aesthetically unpleasant and they interfere with sediment transport and other natural processes, although this can be partly mitigated by artificial BEACH NOURISHMENT. Human removal of beach material continues in some areas today, although legislation has been enacted to discourage it in many areas. The importance of dunes as a natural coastal defence for low-lying land is reflected in laws relating to dune stabilization dating back to the thirteenth century. Humans affect coastal dunes in many direct and indirect ways, including sand extraction, forestation and deforestation, trampling and off-road vehicles, introduction of exotic species and grazing and burrowing animals, and changes in the water table resulting from forestation or residential and industrial development. Dune stabilization and construction has been undertaken in many countries, although it can reduce morphological variety and species diversity. The protection of dunes also impairs their ability to replenish beaches during storms. It has been suggested that construction of a high protective barrier dune on the northern barrier islands of North Carolina threatens their existence, because it prevents OVERWASHING, the opening of inlets and natural barrier recession. Others, however, consider that the artificial dune reduces erosion by nourishing the beach during storms. In many areas, as in dunefields, SALTMARSHes and MANGROVE SWAMPS, one must understand the workings of coastal ecological as well as geomorphological systems to solve coastal problems (Viles and Spencer 1994). Large saltmarsh areas have been reclaimed for agriculture, housing, industry and airports, although there is increasing interest in their preservation with the recognition that they are important and productive ecosystems. Human activities, including deforestation for rice paddies, fuel, construction materials and industrial uses, are continuing to cause irreversible damage to coastal mangroves in tropical regions, however, where there is often little appreciation of their value to native populations. Estuarine dynamics and siltation patterns are being affected by deforestation, mining and quarrying, urbanization, DAM construction, sewage discharge, dredging, dock and marina construction, the reclamation of TIDAL DELTAs, flats and marshes and the diversion of water from one watershed into another. Although much human activity is deleterious to deltas, deforestation, agricultural intensification and extensive soil erosion have sometimes been responsible for their formation or growth. Many deltas are receiving less water and sediment as rivers are dammed for irrigation, flood control and power generation. Much of the loss of wetlands in the Mississippi Delta has natural causes, but it is being exacerbated by deforestation of the drainage basin and dam construction, the building of LEVEEs and other attempts to confine and control the Mississippi River for navigation and flood control. Human activity has been modifying the Nile Delta since predynastic time, but, with construction of the Aswan Dams, almost no fluvial sediment now reaches the delta, and this has resulted in accelerated coastal erosion and marine encroachment. Coral communities and reefs are also threatened by a variety of human activities, including dredging, mining, land clearance, effluents from desalination, sewage discharge, the use of chlorine bleach and explosives for fishing, nuclear weapon testing, oil, chemical and sewerage pollution, thermal pollution from electrical generating stations, careless anchoring, boat grounding and the collection of precious corals and other marine organisms. The Great Barrier Reef Marine Park in Australia was created to manage reefs comprehensively, but economic pressures are more severe in developing areas, and conservation policies more difficult to enforce.

Global warming

One of the greatest challenges facing coastal populations will be to plan for, and manage, the effects of rising sea level resulting from global warming. There is continuing debate over the rate and magnitude of the changes that are to be expected, however, although there has been a trend towards progressively more conservative predictions of sea-level rise in this century. The 2001 third assessment report of a working group for the intergovernmental panel on climatic change (IPCC) has concluded that sea level will rise by between 0.09 and 0.88 m between 1990 and 2100.

Global warming and rising sea level will cause tidal flooding and the intrusion of salt water into rivers, estuaries (see ESTUARY) and groundwater, and it will affect tidal range, oceanic currents, upwelling patterns, salinity levels, biological processes, runoff and landmass erosion patterns. Increasing rates of erosion will make cliffs more susceptible to falls, landslides and other MASS MOVEMENTs, exacerbating problems where loose or weak materials are already experiencing rapid recession. Nevertheless, the effect of rising sea level will vary around the world according to the characteristics of the coast, including its slope, wave climate, tidal regime and susceptibility to erosion.

It has been estimated that about half the world's population lives in vulnerable coastal lowlands, subsiding RIVER DELTAs and river floodplains. The effects of climatic change will be particularly acute in these densely populated regions. It is often the rate of sea-level change rather than the absolute amount that determines whether natural systems, such as coastal marshes and CORAL REEFs, can successfully adapt to changing conditions. Human and natural systems can adjust to slowly changing mean climatic conditions, but it is more difficult to accommodate changes in the occurrence of extreme events. It is not yet known, however, whether higher sea temperatures will increase the frequency and intensity of tropical storms and spread their influence further polewards, or whether higher temperature gradients between land and sea will increase the intensity of monsoons and affect their timing.

Human responses to the rise in sea level will depend upon available resources and the value of the land being threatened. High waterfront values will justify economic expenditure to combat rising sea level in cities, but less attention is likely to be paid to the deleterious effects on saltmarshes, mangroves, coral reefs, lagoons and ice-infested Arctic coasts. The decision-making process associated with coastal erosion and flooding is complex, because of constraints imposed by financial considerations and a myriad of physical, social, economic, legal, political and aesthetic factors. There is public and political pressure on coastal planners and managers to be seen to be doing something about the problem, and this can result in engineering projects that provide only short-term benefits, or which may even exacerbate the original problem. Several managerial options are available, however, ranging from the 'do nothing' approach, to the construction of a completely artificial coast.

References

Davies, J.L. (1972) *Geographical Variation in Coastal Development*, Edinburgh: Oliver and Boyd.

Inman, D.L. and Nordstrom, C.E. (1971) On the tectonic and morphologic classification of coasts, *Journal of Geology* 79, 1–21.

Viles, H. and Spencer, T. (1994) *Coastal Problems*, Oxford: Edward Arnold.

Further reading

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

Carter, R.W.G. and Woodroffe, C.D. (1994) *Coastal Evolution*, Cambridge: Cambridge University Press.

Lakhan, V.C. and Trenhaile, A.S. (eds) (1989) *Applications in Coastal Modelling*, Amsterdam: Elsevier.

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

ALAN TRENHAILE

COHESION

The force by which particles are able to stick together. Cohesion is important in soil mechanics, as it is one of two parameters (alongside the angle of internal friction) that characterize a soil's resistance to an applied stress (though the two parameters are not always independent of each other). Soils with high levels of cohesion (termed cohesive soils) commonly contain a significant amount of clay, which are able to cement the soil internally (yet these typically have low frictional strength). Conversely, dry sand is termed non-cohesive (as particles are easily moved in isolation), with the only resistance to shear coming from the internal friction of sand particles. When sand is moist (though unsaturated) the surface tension of the water menisci between the grains provides an apparent cohesiveness to the sand. This is removed when the sand either dries or becomes saturated. Rocks are commonly high in both parameters. Cohesion becomes proportionately stronger as grain size decreases, allowing fine grain sediments (muds and silts, etc.) to remain stable on high-angle slopes.

Reference

Bullock, M.S., Kemper, W.D. and Nelson, S.D. (1988) Soil cohesion as affected by freezing, water content, time and tillage. *Soil Science Society of America Journal* 52(3), 770–776.

SEE ALSO: adhesion

STEVE WARD

COLLUVIUM

Sedimentary material that has been transported across and deposited on slopes as a result of mass movement processes and soil wash. It is frequently derived from the erosion of weathered bedrock (eluvium) and its deposition on low-angle surfaces, and can be differentiated from material which is deposited primarily by fluvial agency (alluvium). Colluvium can be many metres thick and can infill bedrock depressions (Crozier *et al.* 1990). It often contains palaeosols, which represent halts in deposition, crude bedding downslope, and a large range of grain sizes and fabrics (Bertram *et al.* 1997). Cut-and-fill structures may represent phases when stream incision has been more important than colluvial deposition (Price-Williams *et al.* 1982).

Colluvium may provide a rich record of long-term climatic change (see, for example, Nemec and Kazanci 1999), preserve archaeological materials, indicate phases of accelerated anthropogenic soil erosion during the Holocene and act as a medium into which gullies may be incised (see DONGA).

Colluvial deposits are known from almost all climatic zones from former glacial (Blikra and Nemec 1998) and periglacial environments (Mason and Knox 1997) through to the tropics (Thomas 1994).

References

Bertram, P., Hetu, B., Texier, J.P. and van Steijn, H. (1997) Fabric characteristics of subaerial slope deposits, *Sedimentology* 44, 1–16.

Blikra, L.H. and Nemec, W. (1998) Postglacial colluvium in western Norway: depositional processes, facies and palaeoclimatic record, *Sedimentology* 45, 909–959.

Crozier, M.J., Vaughn, C.E. and Tippett, J.M. (1990) Relative instability of colluvium-filled bedrock depressions, *Earth Surface Processes and Landforms* 15, 329–339.

Mason, J.A. and Knox, J.C. (1997) Age of colluvium indicates accelerated late Wisconsinian hillslope erosion in the Upper Mississippi Valley, *Geology* 25, 267–270.

Nemec, W. and Kazanci, N. (1999) Quaternary colluvium in west-central Anatolia: sedimentary facies and palaeoclimatic significance, *Sedimentology* 46, 139–170.

Price-Williams, D., Watson, A. and Goudie, A. (1982) Quaternary colluvial stratigraphy, archaeological sequences and palaeoenvironment in Swaziland, Southern Africa, *Geological Journal* 148, 50–67.

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

A.S. GOUDIE

COMMINUTION

Refers to the reduction of rock debris to fine powder or to small pieces. In nature, comminution is usually as a result of ABRASION and attrition, and is often linked with problems of coastal erosion due to reduction of shingle.

Further reading

Kabo, M., Goldsmith, W. and Sackman, J.L. (1977) Impact and comminution processes in soft and hard rock, *Rock Mechanics*, Supplementum 9(4), 213–243.

STEVE WARD

COMPACTION OF SOIL

The term compaction refers to a progressive decrease in the volume of a soil element over time, resulting in an increase in density. Recently deposited sediments tend to exhibit a progressive increase in density over time, as consolidation occurs due to self weight and loads imposed by overlying sediment. A commonly used measure of the relative degree of compaction of a soil within engineering soil mechanics is the overconsolidation ratio: OCR = $\sigma'_{max} / \sigma'_{pres}$. Here σ'_{max} refers to the maximum normal EFFECTIVE STRESS which the soil material has experienced over geologic time, while σ'_{pres} is the present-day normal effective stress. Effective stress is defined as total stress minus ambient PORE-WATER PRESSURE (Barnes 2000). Normally consolidated (NC) soils have $\sigma'_{pres} = \sigma'_{max}$, and include most postglacial fluvial and colluvial sediments. Overconsolidated (OC) soils have $\sigma'_{max} \gg \sigma'_{pres}$, and include basal tills and geological strata such as clays and shales which have experienced normal stress reduction caused by erosion of superjacent materials. There are large ranges of OCR from approximately 1.0 to several 100, depending on the history of load changes that the soil has experienced. A transient condition, known as underconsolidation, refers to effective stress below the NC value. This is possible where part or all of the total overburden pressure is borne by the pore fluid, and thus positive excess pore pressures prevail shortly after deposition. It is common where fine-grained, saturated materials are deposited rapidly as QUICKCLAY earthflows or muddy DEBRIS FLOWS. Underconsolidation may also occur where formerly submerged muds become abruptly subaerial, due to either rapid tectonic uplift or lake drainage.

Although the rate of consolidation is controlled strongly by the normal stresses imposed by external loads, soil compaction also varies according to the compressibility of the soil particles themselves, the water content, and the hydraulic conductivity (Barnes 2000). In unsaturated soils, having a high air content, rate of consolidation is controlled primarily by the compressibility of the soil matrix, which is a function of particle shape, sorting, and mineralogy. In saturated soils, rate of consolidation is regulated by soil hydraulic conductivity, since expulsion of virtually incompressible pore fluid is a prerequisite for consolidation. Conductivity varies by several orders of magnitude, depending on particle size and *in situ* density.

Within the normally consolidated class of soils, which comprise many soils worldwide, significant variations in ambient *in situ* density occur as a result of both geomorphic and sedimentological factors. Mixed, poorly sorted materials, such as LANDSLIDE deposits, often possess a relatively high *in situ* density since a wide range of particle sizes ensures that voids between large clasts are filled with finer material (Bement and Selby 1997). It is possible that natural, vibration-induced compaction of rapidly emplaced landslide materials further enhances densification. By contrast, very well-sorted aeolian materials, such as LOESS and DUNE sand, exhibit a much lower *in situ* density, especially if fairly equant grains are dominant in the deposit. Such soils are inherently very compressible.

In the near-surface zone, the effects of geological consolidation are periodically offset by MECHANICAL WEATHERING processes, which lead to a volume increase, and hence a density decrease, relative to that of the unweathered material below. By contrast, the amount of net volume increase brought about by CHEMICAL WEATHERING appears to be slight (Birkeland 1984). In cold regions, FREEZE–THAW CYCLE processes cause seasonal and shorter term cycles of heave and settlement. Thaw and consolidation of the ACTIVE LAYER during spring and summer may produce transient excess pore pressures if the water generated by ice lens melting is slow to escape. This may be due to either a low material conductivity or the existence of an impermeable PERMAFROST table (Williams and Smith 1989). Thaw-consolidation has been credited with the development of very low effective stresses within a thawing active layer, allowing SOLIFLUCTION lobes to move on slope angles as low as $\frac{1}{4}\,\phi'_r$, where ϕ'_r is the residual angle of shearing resistance. Cycles of HYDRATION and dehydration also produce appreciable cyclical volume changes, especially in soils containing montmorillonite clays. However, the magnitudes of the resultant cyclical volume changes are generally far lower than the values attained within seasonally ice-rich sediments.

Rainfall impact, together with infiltration seepage, is also a well-documented soil compacting process, especially in semi-arid environments where it leads to the development of a surface crust of reduced infiltrability. The widespread conversion of grassland and forest soils to arable use has caused significant rainfall compaction of soil, causing reduced infiltrability, and hence accelerated runoff (see RUNOFF GENERATION) and EROSION (Morgan *et al.* 1998). In arable areas, such compaction may be rectified by ploughing and harrowing. In time, uncultivated near-surface soil becomes naturally loosened again by the combined effects of freeze–thaw cycles, bioturbation from soil micro- and macrofauna, in addition to root growth and decay, and downward mixing of low density organic material.

Several problem soils have been identified within engineering soil mechanics based on their poor performance under surcharge stresses or cyclical shear loads. Normally consolidated clays are prone to significant consolidation under structural loads, and may require the placement of fill materials to effect soil consolidation prior to construction (Barnes 2000). NC soils are also more prone to landsliding than are OC materials, since the lesser degree of compaction in the former is generally associated with lower shear strength. A common problem in loess soils is HYDROCOMPACTION (Derbyshire 2001), which involves a localized collapse of soil structure in response to vertical seepage forces. It is a widespread problem where loess is subjected to flood irrigation.

References

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

Bement, R.A.P. and Selby, A.R. (1997) Compaction of granular soils by uniform vibration equivalent to vibrodriving of piles, *Geotechnical and Geological Engineering* 15, 121–143.

Birkeland, P.W. (1984) *Soils and Geomorphology*, Oxford: Oxford University Press.

Derbyshire, E. (2001) Geological hazards in loess terrain, with particular reference to the loess regions of China, *Earth Science Reviews* 54, 231–260.

Morgan, R.P.C., Quinton, J.N. *et al.* (1998) The European soil erosion model (EUROSEM): a dynamic approach for predicting sediment transport from fields and catchments, *Earth Surface Processes and Landforms* 23, 527–544.

Williams, P.J. and Smith, M.W. (1989) *The Frozen Earth*, Cambridge: Cambridge University Press.

MICHAEL J. BOVIS

COMPLEX RESPONSE

Landforms respond to the controlling variables of tectonics, sea level, climate and biotic activity over time. They also respond to the changes, rhythms and thresholds of Earth. Available data suggest that over timescales of 10^2 years increases of geomorphological rates of activity change with a frequency of *c.*2,000 years. Over 10^3 years rates and process balances change with a frequency of 30,000–50,000 years and over 10^{4-5} years full system control changes occur every 100,000–150,000 years. Flux in sediment yield and landform adjustment should be regarded as the norm. Regularity of landform may then be the product of polygenetic landform origins. A central proposition of geomorphology, therefore, is that landform change (response) takes place as states of equilibrium, stability or tranquillity are upset by complex episodic changes to the environmental controls. This may be called 'complex cause' (see LANDSCAPE SENSITIVITY).

The *response* to the hierarchy of controls and events also varies on all timescales and are variably distributed in space. Complex response (Schumm 1973, 1975, 1977, 1979, 1981; Schumm and Parker 1973) describes the way in which the internal structure of a system controls the reaction and relaxation of the system after an impulse of change. There are many aspects to be considered: the effect of internal thresholds (see THRESHOLD, GEOMORPHIC) that control sudden change; the fluctuation between cut-and-fill as the capacity of the system dictates temporary storage of eroded sediment; the effect of area as an impulse moves from a point application (e.g. a river mouth base level change), along a sensitive linear pathway (e.g. a channel, a joint) to diffuse over a catchment as a wave of erosional aggression moving inland (e.g. from a sea cliff or an incising river). Such changes occur after every effective event and the direction of change follows every structural instability.

Landform 'evolution' is a never-ending set of adjustments to impulses of change on all temporal and spatial scales. It is complex.

References

Schumm, S.A. (1973) Geomorphic thresholds and complex responses of drainage systems, in M. Morisawa (ed.) *Fluvial Geomorphology*, Binghamton, Publications in Geomorphology 3, 299–310.

——(1975) Episodic erosion: a modification of the geomorphic cycle, in W.N. Melhorn and R.C. Flemal (eds) *Theories of Landform Development*, 69–86, London: George, Allen and Unwin.

——(1977) *The Fluvial System*, Chichester: Wiley.

Schumm, S.A (1979) Geomorphic thresholds: the concept and its applications, *Transactions Institute of British Geographers*, NS 4, 485–515.
——(1981) Evolution and response of the fluvial system, sedimentological implications, *SEPM, Special Publication*, 31, 19–29.
Schumm, S.A. and Parker, R.S. (1973) Implications of complex response of drainage systems for Quaternary alluvial stratigraphy, *Nature* 243, 99–100.

DENYS BRUNSDEN

COMPLEXITY IN GEOMORPHOLOGY

Complexity is a way of describing complicated, irregular patterns that appear random. It is something tangible that is observable in geomorphic systems, such as in turbulent flow in streams. Much chaotic complexity in geomorphology underlies a larger scale geomorphic order, and overlies smaller scale, more orderly and understandable components. Chaotic turbulent flow is part of a larger scale order seen in the predictable rate and direction of mean streamflow; it is also the result of a huge number of well-understood individual particle trajectories describable by the basic laws of physics. Complexity in geomorphic systems is thus often part of a hierarchy of interrelated structures and processes. Similarly, simple geomorphic patterns, such as beach cusps, commonly arise from complex underlying dynamics; at the same time, they are but a part of broader scale complex patterns. Beach cusps result from complex non-linear interactions between beaches and waves or the complicated formation of edge waves (waves trapped at the shoreline by refraction); at the same time, they are a part of irregular coastline geometry.

One line of explanation for complexity rests in non-linear dynamical systems theory, which has revolutionized many branches of science (see Stewart 1997). To understand the general reasoning involved, it may help to define a few terms first. An unstable system is susceptible of small perturbations and is potentially chaotic. A chaotic system behaves in a complex and pseudo-random manner purely because of the way the system components are interrelated, and not because of forcing by external disturbances, or at least independently of those external factors. The equations describing the system generate the chaos, which is deterministic; chance-like (stochastic) events do not. Systems displaying chaotic behaviour through time usually display spatial chaos, too. Therefore, a landscape that starts with a few small perturbations here and there, if subject to chaotic evolution, displays increasing spatial variability as the perturbations grow. This happens when rivers dissect a landscape and relief increases. Self-organization is the tendency of, for example, flat or irregular beds of sand on streambeds or in deserts to organize themselves into regular spaced forms – ripples and dunes – that are rather similar in size and shape. Self-organization also occurs in patterned ground, beach cusps and river channel networks. Self-destruction (non-self-organization) is the tendency of some systems to consume themselves, as when relief is reduced to a plain. An attractor is a system state that controls system changes and into which other system states are drawn.

Many geomorphic systems are complex, but not all are. Some non-linear geomorphic systems are unstable, chaotic and self-organizing, but some are not. Nevertheless, plentiful evidence suggests that complexity is common in geomorphic systems and begs an explanation. The truly puzzling fact is that most geomorphic systems display order and complexity concurrently. Are the complexities (irregularities) merely deviations from an orderly norm, or are they informative in their own right? A growing body of evidence from field studies, laboratory studies, and real-world datasets suggests that in some geomorphic systems complexity is significant in its own right. Signs of complex behaviour in systems include deterministic chaos, instability, increasing variability over time, self-organization, divergence from similar initial conditions and sensitivity to initial conditions (Phillips 1999: 39–57). Evidence exists for all these indicators of complexity.

Several hydrological records, tree rings series and topographic images reveal chaotic patterns. In other cases, field investigations have confirmed chaotic behaviour predicted in models, as in the genesis of Ultisols in eastern North Carolina.

Field examples of dynamical systems' instability and sensitivity to small perturbations abound, including river meander initiation generated by the unstable growth of small flow perturbations.

Some studies demonstrate patterns of spatial variability that become increasingly complex (less uniform) over time: there is a spatial differentiation of the landscape. Desertification appears to involve an increasingly more complex pattern of vegetation and soil-nutrient resources through time.

In some geomorphic systems, orderly self-organizing patterns seem to emerge from complex non-linear dynamics. Field and laboratory work confirms theoretical work showing that sorted nets in non-periglacial environments may develop spontaneously on any piece of unobstructed land with little or no slope, proving it carries a loose and discontinuous cover of pebbles, each of which may move in small steps with equal probability in all directions (Ahnert 1994).

Much field evidence strongly suggests that some geomorphic systems evolve by diverging from the same, or very similar, initial conditions. In the Norfolk marshes, England, vegetated marsh traps more sediment than bare marsh, so reducing the chances of inundation and lowering (or stabilizing) salinity. The bare marsh becomes lower land that traps more water and the salinity rises, inhibiting vegetation colonization and growth.

Several studies indicate that small variations in initial conditions amplify as a geomorphic system evolves. In podzolized soils in Canada, microtopographic variations produce favoured sites for infiltration and 'funnel' effects that eventually create large variations in the thickness of A and B soil horizons (Price 1994).

Related to complexity are the ideas of fractals and self-organized criticality. Fractal landscapes display self-similar patterns repeated across a range of scales. A small section of coastline may be self-similar to a much larger piece of coastline, of which it is part. Drainage networks, sedimentary layers and joint systems in rocks possess fractal patterns. Self-organized criticality is a theory that systems composed of myriad elements will evolve to a critical state, and that once in this state, tiny perturbations may lead to chain reactions that may affect the entire system. The classic example is a pile of sand. Adding grains one by one to a sandbox causes a pile to start growing, the sides of which become increasingly steep. In time, the slope angle becomes critical: one more grain added to the pile triggers an avalanche that fills up empty areas in the sandbox. After adding sufficient grains, the sandbox overflows. When, on average, the number of sand grains entering the pile equals the number of grains leaving the pile, the sand pile has self-organized into a critical state. Landslides, drainage networks, and the magnitude and frequency relations of earthquakes display self-organized criticality.

Phillips (1999) identifies eleven 'principles of Earth surface systems' that follow from theoretical and empirical work on order and complexity in geomorphic systems. Some of these principles appear to conflict, but that is the nature of complexity. In summary, and applied specifically to geomorphic systems, the principles are (see Huggett 2002: 339–41):

1. Geomorphic systems are inherently unstable, chaotic and self-organizing. Many, but definitely not all, geomorphic systems display a tendency to diverge or to become more differentiated through time in some places and at some times, as when an initially uniform mass of weathered rock or sediment develops distinct horizons.
2. Geomorphic systems are inherently orderly. Deterministic chaos in a geomorphic system is governed by an >attractor = that constrains the possible states of the system. Such a geomorphic system displays dynamic instability but does not behave randomly. The dynamic instability has bounds. Beyond these bounds, orderly patterns emerge that include the chaotic patterns inside them. Thus, even a chaotic system must exhibit order at certain scales or under certain circumstances. For example, at local scales, soil formation is sometimes chaotic, with giant spatial variations in soil properties; as the scale is increased, regular soil–landscape relationships emerge.
3. Order and complexity are emergent properties of geomorphic systems. This principle means that, as the spatial or temporal scale is altered, orderly, regular, stable, and non-chaotic patterns and behaviours and irregular, unstable, and chaotic patterns and behaviours appear and disappear. In debris flows, deterministic chaos governs collisions between particles where the flow is highly sheared and the collisions are sensitive to initial conditions and unpredictable. However, the bulk behaviour of granular flows is orderly and predictable from a relationship between kinetic energy (drop height) and travel length. Therefore, the behaviour of a couple of particles is perfectly predictable from basic physical principles; a collection of particles interacting with each other is chaotic; and the aggregate behaviour of the flow at a still broader scale is again predictable.
4. Geomorphic systems have both self-organizing and non-self-organizing modes. This principle

follows from the first three principles. Some geomorphic systems may operate in self-organizing and non-self-organizing modes at the same time. The evolution of topography, for example, may be self-organizing where relief increases, and self-destructing where relief decreases. Mass wasting denudation is a self-destructing process that homogenizes landscapes by decreasing relief and causing elevations to converge. Dissection is a self-organizing process that increases relief and causes elevations to diverge.

5 Both unstable–chaotic and stable–non-chaotic features may coexist in the same landscape at the same time. Because a geomorphic system may operate in either mode, different locations in the system may display different modes simultaneously. This is the idea of >complex response=, in which different parts of a system respond differently at a given time to the same stimulus. An example is channel incision in headwater tributaries occurring concurrently with valley aggradation in trunk streams.

6 Simultaneous order and disorder, observed in real landscapes, may be explained by a view of Earth surface systems as complex non-linear dynamical systems. They may also arise from stochastic forcings and environmental processes.

7 The tendency of small perturbations to persist and grow over finite times and spaces is an inevitable outcome of geomorphic system dynamics. In other words, small changes are sometimes self-reinforcing and lead to big changes. Examples are the growth of nivation hollows and dolines. An understanding of non-linear dynamics helps to determine the circumstances under which some small changes grow and others do not.

8 Geomorphic systems do not necessarily evolve towards increasing complexity. This principle arises from the previous principles and particularly from Principle 4. Geomorphic systems may become more complex or simpler at any given scale, and may do either at a given time.

9 Neither stable, self-destructing nor unstable, self-organizing evolutionary pathways can continue indefinitely in geomorphic systems. No geomorphic system changes ad infinitum. Stable development implies convergence that eventually leads to a lack of differentiation in space or time, as when different elevations in a landscape converge to form a plain. Disturbances disrupt such stable states by reconfiguring the system and resetting the geomorphic clock. Divergent evolution is also self-limiting. For example, base levels ultimately limit landscape dissection.

10 Environmental processes and controls operating at distinctly different spatial and temporal scales are independent. For example, processes of wind transport are effectively independent of tectonic processes, although there are surely remote links between them.

11 Scale independence is a function of the relative rates, frequencies and durations of geomorphic phenomena.

References

Ahnert, F. (1994) Modelling the development of non-periglacial sorted nets, *Catena* 23, 43–63.

Huggett, R.J. (2002) *Fundamentals of Geomorphology*, London: Routledge.

Phillips, J.D. (1999) *Earth Surface Systems: Complexity, Order, and Scale*, Oxford: Blackwell.

Price, A.G. (1994) Measurement and variability of physical properties and soil water distribution in a forest podzol, *Journal of Hydrology* 161, 347–364.

Stewart, I. (1997) *Does God Play Dice? The New Mathematics of Chaos*, new edn, Harmondsworth: Penguin Books.

Further reading

Culling, W.E.H. (1988) A new view of the landscape, *Transactions of the Institute of British Geographers* New Series 13, 345–360.

Hergarten, S. and Neugebauer, H.J. (2001) Self-organized critical drainage, *Physical Review Letters* 86, 2,689–2,692.

Phillips, J.D. (1999) Divergence, convergence, and self-organization in landscapes, *Annals of the Association of American Geographers* 89, 466–488.

——(2000) Signatures of divergence and self-organization in soils and weathering profiles, *Journal of Geology* 108, 91–102.

Richards, A.E. (2002a) Complexity in physical geography, *Geography* 87, 99–107.

Richards, A.E., Phipps, P. and Lucas, N. (2000) Possible evidence for underlying non-linear dynamics in steep-faced glaciodeltaic progradational successions, *Earth Surface Processes and Landforms* 25, 1,181–1,200.

Xu, T., Moore, I.D., and Gallant, J.C. (1993) Fractals, fractal dimensions and landscapes – a review, *Geomorphology* 8, 245–262.

RICHARD HUGGETT

COMPUTATIONAL FLUID DYNAMICS (CFD)

Fluid motions play a central role in sculpting a great variety of landforms, both terrestrial and submarine. Examples range from river channels to aeolian dunes to barrier islands. Naturally, investigations into landform origins often involve applications of fluid dynamics. Fluid dynamics is that branch of mechanics that concerns the physics of fluid motion. The motion of a non-turbulent, Newtonian fluid is described by the Navier–Stokes equations. These equations express continuity of mass and momentum in three dimensions in a continuum fluid subject to gravitational, inertial, viscous and pressure forces. The equations take on different forms depending on whether the fluid is compressible (e.g. air) or incompressible (e.g. water to a close approximation). Normally the Navier–Stokes equations are combined with models of turbulence (for application to turbulent flows) and with models of boundary friction (for any flows involving contact with a surface, such as a channel bed). Except in special cases, the Navier–Stokes equations cannot be analytically solved. Their solution can, however, be approximated using numerical methods (see, e.g. Cheney and Kincaid 1999; Press *et al.* 1993) combined with a set of specified initial and boundary conditions. Such methods involve dividing up space and time into discrete elements, within which the variables of interest – such as velocity and pressure – are either interpolated or held constant. The development of numerical solution methods for different types of equations, including fluid flow equations, is a major area of research in the fields of mathematics and computing science. Numerical solutions of equations for fluid motion can be quite computationally intensive, and the computer models that implement these solutions are referred to as Computational Fluid Dynamics (CFD) models. Depending on the methods used and the degree of approximation involved, the computer codes can be quite complex, and there are many commercially available packages as well as research codes developed within universities. Applications of CFD are increasingly widespread in geomorphology. CFD models have been of great benefit, for example in understanding interactions between fluid flow, bed morphology and sediment transport in river channels (e.g. Hankin *et al.* 2002; Lane *et al.* 2002; Ma *et al.* 2002). CFD models of airflow dynamics have been used to understand the interactions between airflow and dune morphology (e.g. Walmsley-John and Howard 1985). Other applications have been wide-ranging; examples include water flow in karst conduits (e.g. Hauns *et al.* 2001), circulation and sediment movement in ancient epeiric seas (e.g. Slingerland *et al.* 1996), coastal morphology (e.g. Deigaard and Fredsoe 2001) and paleoflood hydrology (e.g. House and Baker 2001).

References

Cheney, W. and Kincaid, D. (1999) *Numerical Mathematics and Computing*, 4th edition, Pacific Grove, CA: Brooks/Cole.

Deigaard, R. and Fredsoe, J. (2001) The use of numerical models in coastal hydrodynamics and morphology, in G. Seminara, and P. Blondeaux (eds) *River, Coastal and Estuarine Morphodynamics*, 61–92, Berlin: Springer-Verlag.

Hankin, B.G., Holland, M.J., Beven, K.J. and Carling, P. (2002) Computational fluid dynamics modelling of flow and energy fluxes for a natural fluvial dead zone, *Journal of Hydraulic Research* 40(4), 389–402.

Hauns, M., Jeannin, P.-Y. and Atteia, O. (2001) Dispersion, retardation, and scale effect in tracer breakthrough curves in karst conduits, *Journal of Hydrology* 241(3–4), 177–193.

House, P.K. and Baker, V.R. (2001) Paleohydrology of flash floods in small desert watersheds in western Arizona, *Water Resources Research* 37(6), 1,825–1,839.

Lane, S.N., Hardy, R.J., Elliot, L. and Ingham, D.B. (2002) High-resolution numerical modelling of three-dimensional flows over complex river bed topography, *Hydrological Processes* 16(11), 2,261–2,272.

Ma, L., Ashworth, P.J., Best, J.L., Elliot, L., Ingham, D.B. and Whitcombe, L.J. (2002) Computational fluid dynamics and the physical modelling of an upland urban river, *Geomorphology* 44(3–4), 375–391.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1993) *Numerical Recipes in C*, 2nd edition, Cambridge: Cambridge University Press.

Slingerland, R., Kump, L.R., Arthur, M., Fawcett, P., Sageman, B. and Barron, E. (1996) Estuarine circulation in the Turonian Western Interior Seaway of North America, *Geological Society of America Bulletin* 108, 941–952.

Walmsley-John, L. and Howard, Alan D. (1985) Application of a boundary-layer model to flow over an eolian dune, *Journal of Geophysical Research, D, Atmospheres* 90(6), 10,631–10,640.

Further reading

Feynman, R.P., Leighton, R.B. and Sands, M.L. (1989) *The Feynman Lectures on Physics: Commemorative Issue* (3 volume set), Redwood City, CA: Addison-Wesley.

White, F.M. (1998) *Fluid Mechanics*, 4th edition, Boston, MA: McGraw-Hill.

GREG TUCKER

CONCHOIDAL FRACTURE

A smoothly curved fracture, marked by concentric rings and resembling a bi-valve shell in shape. Conchoidal fractures are the most common type of fracture, and are also known as clamshell fractures. They occur when bonds between atoms are approximately the same in all directions within a mineral, and result in breakage along smooth, curved surfaces. Conchoidal fractures occur particularly in amorphous materials (i.e. those showing no definite crystalline structure) such as obsidian, and are also common in quartz, chert and glass.

Further reading

Atkinson, B.K. (1987) *Fracture Mechanics of Rock*, London: Academic Geology Series, Academic Press.

STEVE WARD

CONFLUENCE, CHANNEL AND RIVER JUNCTION

River channel confluences, the sites at which two open channels combine, are ubiquitous features of all river networks and channel patterns. These sites mark nodes of significant change in hydraulic geometry (Richards 1980), flow and sediment discharge, and are characterized by a complex three-dimensional flow field and variable bed geometry (Mosley 1976; Best 1988; Bradbrook *et al.* 2000; Rhoads and Sukhodolov 2001). River channel junctions are often points of significant bed scour (e.g. Best and Ashworth 1997), and are critical in considerations of sediment/pollutant dispersal and mixing in channel networks (Plate 27). Study of these complex fluvial sites has progressed through field, physical and numerical modelling and has identified five principal controls on flow, sediment transport and bed morphology at channel confluences: (1) the angle of convergence between the confluent channels; (2) the ratio of discharge, or flow momentum, between the incoming channels; (3) the planform shapes of the junction (for instance, 'Y' or '⊥' shaped) and upstream channels (i.e. curved, straight; single, multiple); (4) the presence of any depth differential between the incoming channels; and (5) the relative roughness of the confluence (ratio of flow depth to grain size), with the hydrodynamic influence of the particles beginning to dominate at larger grain sizes (e.g. Roy *et al.* 1988).

Plate 27 Junction of the Paraná and Paraguay Rivers, Argentina. The Paraguay River enters from the right and is picked out by its higher suspended sediment concentration. The shear layer between the two rivers displays a series of large vortices and the mixing layer remains distinct for many tens of kilometres downstream. Width of Paraguay River inflow at confluence ~1 km

Fluid dynamics

Channel confluences are zones of complex, three-dimensional, turbulent flow where significant local flow acceleration and deceleration may occur due to both the increased combined fluid discharge and the specific fluid dynamics of the confluence region. Experimental, field and numerical studies have shown confluences to be dominated by seven fluid dynamic zones (Figure 26).

1 A zone of flow stagnation near the upstream junction corner; this fluid deceleration is caused by turning and hence centrifugal forcing of the flows as they approach the junction, together with the influence of a pressure gradient within the junction that is generated by water surface superelevation in the junction centre.

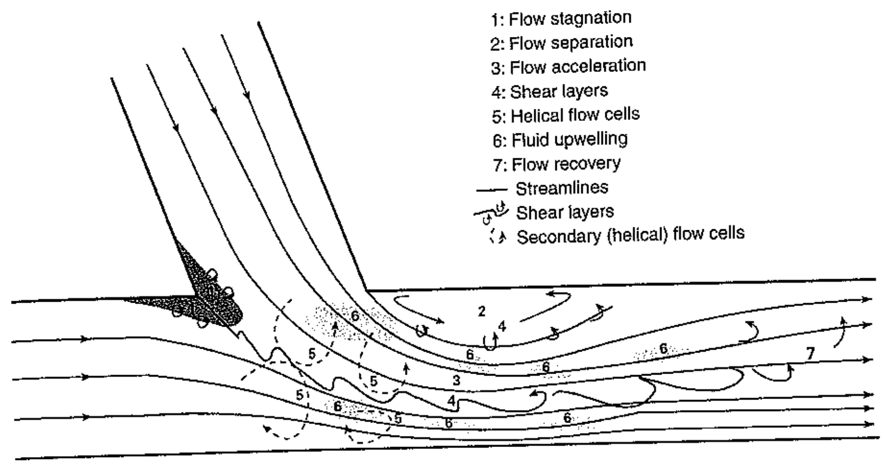

Figure 26 Schematic diagram of the seven principal fluid dynamic zones that may be present at channel confluences

2 A region of flow separation may occur downstream from the downstream junction corner(s); flow cannot remain attached to the boundary at sudden changes in geometry, and an adverse hydrostatic pressure gradient here causes the flow to separate from the wall and form a region of slow, recirculating flow. In symmetrical confluences, the downstream separation zone may form on both sides of the junction. In an asymmetric confluence, the downstream separation zone may only form on the angled (i.e. tributary) side of the junction. The size of the downstream separation zone(s) increase(s) with junction angle and tributary discharge (Best 1988; Bradbrook *et al.* 2000), but may be modified/absent at natural junctions where the angle of bank divergence at the downstream junction corner(s) may be modified by formation of a point bar through sediment deposition (e.g. Rhoads and Sukhodolov 2001) and/or bank erosion.

3 A region of flow acceleration forms at the centre of the confluence that is generated by both the increased fluid discharge passing through the junction (see streamline convergence, Figure 26) and also the constricting influence of any region of flow separation.

4 Distinct shear layers are generated along regions where velocity gradients are severe. Thus, shear layers can be present on either side of the flow stagnation region, along the mixing interface between the two joining flows, bounding any regions of flow separation and also arising from any steep changes in bed topography (such as the avalanche faces that may dip into the central scour – see below). Large, turbulent and 3D flow structures arising along these shear layers, termed Kelvin–Helmholtz instabilities, may give rise to high turbulent shear stresses that are influential in fluid mixing and sediment transport.

5 Helical flow may develop within the junction due to the presence of streamline curvature (Figure 26; streamlines are lines drawn in the fluid of which the tangent at any point is the direction of velocity at that point) and water surface superelevation within the confluence. In ideal cases, where the tributaries are near symmetrical and have equal flow momentum, these secondary flows may be expressed as surface convergent, bed divergent flows much as in placing two meanders back to back, although the duration of streamline curvature through the bend means that it is unlikely that an entire helix is ever completed. The presence of both flow separation at the junction corner, changing pressure gradients or flow separation associated with bed

topography ('topographic forcing' of the flow) or a depth differential between the two incoming tributaries, may lessen the effects or destroy such large-scale secondary flows. Additionally, the time-averaged picture of a series of individual turbulent events, such as fluid upwelling in the confluence, may be manifested as apparent secondary circulation (Lane *et al.* 2000).

6 Regions of distinct fluid upwelling may be generated by both distortion of the shear layer and flow associated with bed topography, such as where the beds of the tributaries are discordant in their height at the junction. This may encourage upwelling of one stream into the other, thus greatly increasing the rate of mixing at the junction (Gaudet and Roy 1995).

7 Finally, a region of flow recovery downstream of the confluence has been observed. This is where the effects of the junction lessen and flow returns to a more uniform cross-stream distribution. However, the flows may remain unmixed for many channel widths downstream if the velocity differential across the shear layer is minimal and the local turbulence at the junction does not mix the flows (see Plate 27).

Bed morphology

The topography of river channel confluences is often characterized by four distinct elements. First, a central scour hole is often present whose orientation approximately bisects the junction angle. The depth of scour increases at both higher junction angles and momentum ratios, and some of the largest alluvial scours are found at these sites. Scours at junctions may reach between two and ten times the depth of the upstream confluent channels: for instance, scour depth at the confluence between the Ganges and Jamuna (Brahmaputra) Rivers in Bangladesh has been recorded as up to 30 m below the upstream bed level in the confluent channels (Best and Ashworth 1997). The position and cause of the confluence scour have been related to (a) flow acceleration in the centre of the confluence (e.g. Roy *et al.* 1988); (b) the influence of turbulence along the shear layer between the flows; (c) downwelling created by secondary flows that may cause higher momentum fluid to be transferred towards the bed at the centre of the confluence; and (d) the differential routing of sediment around the scour.

Second, tributary mouth bars have been observed that terminate at the junction. These bars may possess a steep slipface that dips into the central scour, although the angle of this surface can range from only a few degrees to angle-of-repose for the sediment (~20°–35°). The position of these faces is controlled by the momentum ratio between the confluent channels, with tributary mouth bars migrating further into the junction as the discharge from that channel becomes a greater fraction of the combined confluence flow.

Third, bars may form within regions of flow separation downstream of the junction corners. Flow separation provides a low velocity region into which sediment can accumulate and these bars may show appreciable fining of sediment since only the finer grained sediment can be entrained into this area. Accumulation of sediment in this region will alter the velocity and pressure gradients within this zone and may lead to a lessening of the extent and influence of flow separation.

Finally, mid-channel bars may form in the region of flow deceleration downstream of the junction scour, especially in 'Y'-shaped junctions, or where sediment delivery is high, and they may mark regions of deposition of sediment eroded at the junction scour. Ferguson (1993) has identified the confluence–diffluence unit as a fundamental braided river building block, in which confluence scour creates the sediment that, as the channel widens to cope with the increased discharge, encourages mid-channel bar development and diffluence formation. However, little study has been conducted on sediment transport through confluences, although experimental work suggests that the bed scour may be a zone of reduced transport rates and that sediment may be routed around and not through the scour. This also reflects the streamline pattern within the junction (Figure 26) and the influence of both shear layers and secondary flows within the confluence.

References

Best, J.L. (1988) Sediment transport and bed morphology at river channel confluences, *Sedimentology* 35, 481–498.

Best, J.L. and Ashworth, P.J. (1997) Scour in large braided rivers and the recognition of sequence stratigraphic boundaries, *Nature* 387, 275–277.

Bradbrook, K.F., Lane, S.N. and Richards, K.S. (2000) Numerical simulation of three-dimensional, time-averaged flow structure at river channel confluences, *Water Resources Research* 36, 2,731–2,746.

Ferguson, R.I. (1993) Understanding braiding processes in gravel-bed rivers: progress and unresolved problems, in J.L. Best and C.S. Bristow (eds) *Braided Rivers*, Geological Society of London Special Publication 75, 73–87.

Gaudet, J.M. and Roy, A.G. (1995) Effect of bed morphology on flow mixing length at river confluences, *Nature* 373, 138–139.

Lane, S.N., Bradbrook, K.F., Richards, K.S., Biron, P.M. and Roy, A.G. (2000) Secondary circulation cells in river channel confluences: measurement artefacts or coherent flow structures?, *Hydrological Processes* 14, 2,047–2,071.

Mosley, M.P. (1976) An experimental study of river channel confluences, *Journal of Geology* 84, 535–561.

Rhoads, B.L. and Sukhodolov, A.N. (2001) Field investigation of three-dimensional flow structure at stream confluences: 1: Thermal mixing and time-averaged velocities, *Water Resources Research* 37, 2,393–2,410.

Richards, K.S. (1980) A note on change in geometry at tributary junctions, *Water Resources Research* 16, 241–244.

Roy, A.G., Roy, R. and Bergeron, N. (1988) Hydraulic geometry and changes in flow velocity at a river confluence with coarse bed material, *Earth Surface Processes and Landforms* 13, 583–598.

Further reading

Best, J.L. and Roy, A.G. (1991) Mixing layer distortion at the confluence of channels of different depth, *Nature* 350, 411–413.

Biron, P., Roy, A.G. and Best, J.L. (1996) Turbulent flow structure at concordant and discordant open-channel confluences, *Experiments in Fluids* 21, 437–446.

JIM BEST AND STUART LANE

CONTINENTAL SHELF

The continental shelf generally is defined as the zone adjacent to a continent or around an island that is between the shoreline and a noticeable break in slope, the shelf break, to the steeper continental slope or, where there is no break in slope, to a depth of about 200 m. Along with the coastal plain, continental slope and continental rise, the shelf is considered part of the continental margin (Gary *et al.* 1972: 153) and usually is synonymous with the term continental platform (Baker *et al.* 1966: 38). The division of the shelf into the inner, mid (occasionally) and outer continental shelf is arbitrary and may be based on logical or scientific criteria, such as the depth to which waves agitate the seafloor, or by legal criteria, such as the geographic limit of jurisdiction by a government. In many regions, the continental shelf is physically continuous with the coastal plain; the separation between the two being the location of the shoreline. The dynamic nature of the continental shelf is symbolized by the active character of the shoreline which moves laterally and vertically in spatial and temporal scales that vary by orders of magnitude.

The shelf is important for several reasons. It is a zone of many physical and biological transitions from oceanic to terrestrial conditions and processes. As everything that moves from land to the ocean must pass across the continental shelf, the suite of processes acting on the shelf is vitally important. The shelves are regions of abundant biological activity as there are substantial supplies of nutrients from both upwelling and upland runoff and there generally is good light penetration. Finally, the continental shelves are sites of major economic interest ranging from the commercial and recreational fisheries of the shelf waters to the sands, gravels and other hard minerals of the surficial sediments to oil and gas that have formed from included biotic sediments and accumulated in any of several types of traps within the body of the shelf.

At the smallest scales, the shoreline and, hence, the boundary between the coastal plain and continental shelf shifts within seconds and hours in response to waves and tides. While, probably more importantly, the multi-millennial, glacial-eustatic SEA-LEVEL changes during the Quaternary have moved the shoreline several tens of kilometres laterally and a hundred or so metres vertically. Furthermore, consequences of local or regional tectonic activity, which usually is spasmodic, are additive to eustatic trends. The presence, or absence, and cause of the tectonics contribute to the overall form of the shelf. The proximity of the continental shelf to the edge of a crustal plate (see PLATE TECTONICS) and the type of inter-plate dynamic play crucial roles in the form and function of the shelf.

Perhaps the least geologically mature shelves are those along convergent plate boundaries and other ACTIVE MARGINS as commonly occur around much of the Pacific Ocean and along the northern shore of the Mediterranean Sea. Although this situation presents a potentially complex and geologically interesting continental margin, the rate of tectonic activity tends to limit the length of time during which marine processes are able to act on a specific body of sediment or location on the continental shelf. However, the same processes that restrict the geographic domain of the shelf result in a rapid, gravity driven flux of material between the often steep and high, near-coastal continental areas and the deep ocean. Milliman and Syvitski (1992) indicate that the small drainage, high relief river systems of active margins contribute a vast quantity of sediment to the ocean basins. Residence time of sediment on the shelf is short and the movement of the sediment across the narrow continental shelf mostly is controlled by oceanographic processes that respond to shelf morphology among other factors. As an example, the zone in which WAVES shoal and resuspend bottom sediments is relatively narrow. This narrowness, in turn, results in sharp gradients in the intensity of wave transformation and related processes.

The contrasting situation is a continental shelf on a PASSIVE MARGIN well removed from a spreading centre, as is the situation along much of the Atlantic coasts of North and South America, Europe and Africa. Such broad, gently sloping continental margins can be significant sites of sediment accumulation over an extensive time. Studies along the east coast of North America indicate a kilometres-thick depositional sequence that began with the filling of early Mesozoic rift valleys (see RIFT VALLEY AND RIFTING) or basins and continues through the present.

Large-scale – many tens of metres – changes in sea level play a major role in the development of the continental shelf. Wright (1995), studying the mid-Atlantic shelf of North America, considers the cumulative time during which any portion of the seafloor potentially is subject to wave energy of sufficient magnitude to agitate the bottom sediments. This zone extends from the shoreline/surf zone offshore to a depth determined wave dynamics and assumptions about the likelihood of specific waves occurring within the area. The width of this zone of bottom agitation primarily is a function of the slope of the shelf surface. The rate of movement of the zone across the shelf is a function of both the rate of sea-level change and the bottom slope. In regions such as that studied by Wright (1995), where sea-level history mainly is governed by eustasy, the determination of the duration of potential bottom activity is comparatively straightforward whereas in areas with a complex history, with major tectonic or glacio-isostatic sea-level component (Kelley *et al.* 1992), the process history is more complex.

In addition to growth by upward or outward sedimentation, other factors can influence the trapping of sediments and the subsequent form of the shelf. Lengthy barrier reefs, shore parallel lines of DIAPIRS, fault blocks or folds can form dams to cross-shelf sediment transport with consequent ponding of sediments. In the situation where the offshore shelf dam has substantial relief and catches a significant quantity of sediment, the mass of the accumulated sediments can trigger isostatic subsidence which results in a deepening of the depositional basin and further trapping of sediment. This seems to have been the case with the growth of the up to 15-km thick Baltimore Canyon Trough which appears to have formed both as the fill in Mesozoic grabens or rift basins and, in places, behind a Jurassic/Cretaceous barrier reef (Schlee 1980).

Several factors including shelf width and slope, rate of change of relative sea level, the availability of sediment and the characteristics of that sediment, and the intensity of physical oceanographic processes determine whether a continental shelf builds laterally or vertically or does not accrete while serving as a conduit for sediment moving from the continent to the deep sea. Similarly, the interaction of the rate and locus of sediment deposition on the shelf with the rate of sea-level rise influences whether an area experiences marine transgression or regression. An understanding of the occurrence and forms of RIVER DELTAS may serve as a surrogate for a similar knowledge of continental shelf growth especially as most deltas grow on or across the shelf. These same factors in combination with others, such as climate, determine the character of sediments that are resident on and within the shelf. Hayes (1967) observed that mud is a major constituent of inner continental shelf sediments offshore of areas with high temperature and high rainfall (strong CHEMICAL WEATHERING), coral is most common in areas with high temperatures, gravel is most common offshore of areas with low temperatures (where MECHANICAL WEATHERING dominates and there is substantial ice transport of large particles), and that rock is abundant in cold areas (perhaps due to scouring of sediments by ice) but is strongly correlated with the slope of the inner shelf.

References

Baker, B.B., Jr, Deebel, W.R. and Geisenderfer, R.D. (eds) (1966) *Glossary of Oceanographic Terms*, Washington, DC: US Naval Oceanographic Office.

Gary, M., McAffee, R., Jr and Wolf, C.L. (eds) (1972) *Glossary of Geology*, Washington, DC: American Geological Institute.

Hayes, M.O. (1967) Relationship between climate and bottom sediment type on the inner continental shelf, *Marine Geology* 5, 111–132.

Kelley, J.T., Dickson, S.M., Belknap, D.F. and Stuckenrath, R., Jr (1992) Sea-level change and Late Quaternary sediment accumulation on the Southern Maine inner continental shelf, in C.H. Fletcher, III and J.F. Wehmiller (eds) *Quaternary Coasts of the United States: Marine and Lacustrine Systems*, Tulsa, OK: SEPM (Society of Sedimentary Geology).

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

Schlee, J.S. (1980) Seismic stratigraphy of the Baltimore Canyon Trough, *US Geological Survey Open File Report* 80–1,079.

Wright, L.D. (1995) *Morphodynamics of Inner Continental Shelves*, Boca Raton, FL: CRC Press.

CARL H. HOBBS, III

CONTRIBUTING AREA

In hydrological terms a contributing area is the part of a DRAINAGE BASIN that provides stormwater RUNOFF GENERATION. The link between precipitation input and catchment outflow is largely determined by variability in soil moisture storage and the spatial distribution of contributing areas for surface runoff. Almost all stormwater runoff is generated by surface or near-surface flow processes. Therefore runoff-contributing areas within drainage basins are mainly dominated by subsurface stormflow and OVERLAND FLOW. Two processes can generate overland flow. Infiltration-excess overland flow, occurs when precipitation intensity exceeds the rate of water infiltration into the soil. This process tends to occur in catchments in semi-arid regions where natural vegetation is sparse or where there has been disturbance of the land (e.g. extensive agriculture). The second process is saturation-excess overland flow, which occurs when precipitation falls on a saturated soil surface. During a storm, when antecedent soil-moisture conditions in a catchment are high, the water table may temporarily intersect with the ground surface producing saturation-excess overland flow.

The spatial extent and pattern of runoff-contributing areas is affected by climate, soil and topography. Contributing areas of infiltration-excess overland flow are determined by the interaction of rainfall intensity and soil permeability. The least permeable soils in a basin are the most likely to contribute infiltration-excess overland flow. As rainfall intensity increases, areas with more moderate permeability also may contribute overland flow.

However, at the start of rainfall soil moisture is not evenly distributed but is concentrated in the areas adjacent to perennial water courses and in topographic hollows. Overland flow may be generated by return flow when seepage is concentrated and surface soils become fully saturated. Under these circumstances the water table is high and ground water is in close proximity to the surface. These areas preferentially generate storm runoff so the storm hydrograph peak is generated from a relatively small part of the catchment – the partial contributing area (Betson 1964). This runoff-producing area will expand during the course of a storm.

Figure 27 shows the extent of saturation in a small drainage basin at three stages: pre-storm, mid-storm and late storm. Prior to a storm the area of saturation is preferentially concentrated in hollows and in soils adjacent to stream channels. As the storm progresses the saturated area expands into the hillslope hollows at the channel heads creating saturated overland flow from return flow. This coalesces into stream flow resulting in extension of the channel network. By late storm, channel heads are fully saturated and small perennial streams are flowing. The question as to where channels begin has been addressed in a model by Montgomery and Dietrich (1988) who predict the contributing area required for channel initiation in channel heads generated in landslide hollows. It follows that the areas contributing to runoff in a drainage basin are fairly restricted, occurring mainly at the base of slopes or channel heads where subsurface runoff is at its maximum and groundwater tables are very shallow; where subsurface flow converges in the soil in hillslope hollows; and areas of reduced soil moisture storage.

The importance of the contributing area idea is underpinned by several important hydrological concepts. Betson (1964) developed the concept of partial area storm runoff. This was based on a series of simple mathematical models that used Hortonian infiltration theory to predict the areas contributing to runoff during a storm. The

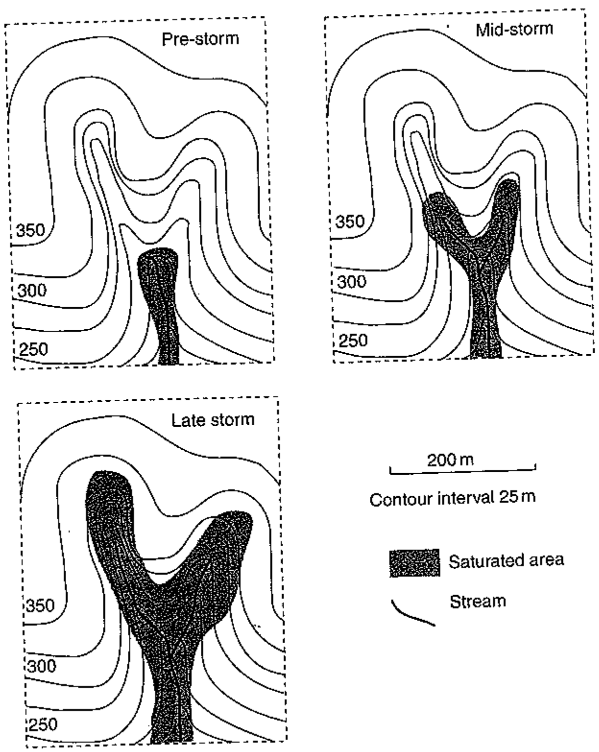

Figure 27 Sequence of expansion of the saturated area of a first-order stream catchment in response to a storm event

equations developed, which can be thought of as functions of apparent watershed infiltration capacity, demonstrate that runoff originates from a small, but relatively consistent, part of the catchment. Using the basic hydrological variables of storm precipitation, storm duration and runoff volume, Betson defined the contributing area as peak stream runoff divided by peak rainfall intensity. This ratio defines the effective runoff-producing area of a drainage basin, which can be expressed as a percentage or proportion of the total catchment area. This is calculated over a storm as total storm runoff divided by total storm rainfall. Typical values are less than 10 per cent for small well-vegetated catchments. The observation that storm runoff frequently occurs from only a small part of the catchment and the size of the runoff-contributing area does not vary greatly within a drainage basin is the basis of the partial area storm runoff concept. However, because this idea is based on Hortonian infiltration-excess runoff theory this is not generally applicable to all catchments.

Working in the humid forests, workers began to recognize the importance of subsurface stormflow generation by throughflow and saturated overland flow at rainfall intensities far less than infiltration excess overland flow (Troendle 1985). This led to the concept of the variable source area, whereby storm runoff was generated in only certain parts of the catchment (Hewlett 1961). The extent of the runoff-generating areas varied from storm to storm and from season to season. On lower slope where groundwater levels are nearest the surface and soil water seepage results in elevated soil moisture storage during a storm, subsurface flow may resurge towards the base as saturated overland flow. Hewlett (1961) and Hewlett and Hibbert (1967), working in forested catchments of North Carolina, demonstrated the importance of this runoff mechanism as opposed to infiltration-excess overland flow so widely popularized by Horton. Other work, particularly by Dunne and Black (1970) working in Vermont, USA clearly established saturation overland flow could be the dominant source of stormwater runoff in a stream.

These ideas are manifest today in the partial contributing area concept which is implicit in the dynamic contributing area concept in recognition of the fact that the area contributing runoff is not fixed but expands during a storm as the saturated areas at the foot of slopes and channel heads extend. When precipitation stops and slopes begin to drain the contributing areas contract. Given that contributing areas are defined by the spatial pattern of surface storm runoff, including overland flow, topography is fundamental in determining the extent. For example, hillslope hollows and swales tend to concentrate saturated overland flow.

Contributing areas of saturation-excess overland flow are determined by the interaction of topography and soil-moisture conditions (Anderson and Burt 1978). The degree of concentration is determined by the area drained per unit contour length (a) and the local slope gradient (s). The a/s index (Kirkby 1978) defines areas of flow convergence and divergence that dictate local drainage conditions for both saturated overland flow and seepage. This topographic control on saturation-excess overland flow can be quantified for the drainage basin as a whole using the topographic wetness index (TWI) (Wolock and McCabe 1995). The TWI is calculated as $\ln(a/s)$ for all points in a catchment. The areas of a catchment with the highest TWI values are the most likely to contribute saturation-excess overland flow. During dry periods when soil-moisture storage is low, only areas with the very highest TWI values are likely to be saturated and contribute overland flow runoff. Under saturated conditions areas with lower TWI values will contribute to runoff.

Land use strongly affects the nature of runoff within a catchment, both in terms of physical processes and solute dynamics. Factors such as surface vegetation, soil permeability and land management practices determine the relative

importance of runoff from different types of land use. Furthermore, the pathway of flow through the soil is likely to alter the solute balance of stormwater runoff (e.g. the take-up of nitrates from agricultural fertilizer). In this respect the land use not only influences runoff pathways but will also be important in controlling the sources, types and amounts of contaminants that enter runoff. Furthermore, the dominance of surface or near-surface flow processes in generating storm runoff is an important consideration in the stability of slopes. Many soil mechanics problems can only be addressed by having a good knowledge of hillslope hydrology.

The concept of contributing areas within drainage basins has provided much better understanding of stormwater runoff mechanisms. This has led to better hydrological predictions and the development of distributed runoff models. These models can now be coupled with sediment transport and erosion models to provide realistic simulations of drainage basin development (Willgoose *et al.* 1991).

References

Anderson, M.G. and Burt, T.P. (1978) The role of topography in controlling throughflow generation, *Earth Surface Processes and Landforms* 3, 331–344.

Betson, R.P. (1964) What is watershed runoff? *Journal of Geophysical Research* 69, 1,541–1,551.

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

Hewlett, J.D. (1961) Some ideas about storm runoff and baseflow, *United States Department of Agriculture, Forest Service Southeast Forest and Range Experiment Station, Annual Report*, 62–6.

Hewlett, J.D. and Hibbert, A.R. (1967) Factors affecting the response of small watersheds to precipitation in humid areas, in *Proceedings of the International Symposium on Forest Hydrology*, 275–290, Oxford: Pergamon.

Kirkby, M.J. (1978) *Hillslope Hydrology*, Chichester: Wiley.

Montgomery, D.R. and Dietrich, W.E. (1988) Where do channels begin? *Nature* 336, 232–234.

Troendle, C.A. (1985) Variable source are models, in M.G. Anderson and T.P. Burt (eds) *Hydrological Forecasting*, 347–403, Chichester: Wiley.

Willgoose G.R., Bras R.L. and Rodriguez-Iturbe, I. (1991) Results from a new model of river basin evolution, *Earth Surface Processes and Landforms* 16, 237–254.

Wolock, D.M. and McCabe, G.J. (1995) Comparison of single and multiple flow-direction algorithms for computing topographic parameters in TOPMODEL, *Water Resources Research* 31, 1,315–1,324.

SEE ALSO: drainage basin; models; overland flow; runoff generation

JEFF WARBURTON

CORAL REEF

Coral reefs are natural structures of calcium carbonate made largely from the skeletons of hard corals and coralline algae. Some modern reefs have been forming for millions of years and can stretch for hundreds of kilometres off tropical coasts.

Distribution in time and space

Coral reefs are found mainly between 25°N and 25°S. The reef-building (hermatypic) corals and associated organisms live best in sea-surface temperatures between 25 °C and 29 °C. Hermatypic corals mostly live only within the upper few metres of water, the 'photic' zone into which sufficient light can penetrate for their symbiotic algae (zooxanthellae) to be able to photosynthesize.

In a general sense, the distribution of fossil coral reefs suggest that sea-surface temperatures have constrained their spread since their appearance in the early Triassic (Birkeland 1997). At a sub-regional scale, other factors were important. For example, the presence of terranes in the central tropical Pacific aided the dispersal of corals across the wider-than-present Pacific during the Palaeozoic and much of the Mesozoic (Grigg and Hey 1992). West–east ocean currents helped the development of coral reefs in the easternmost Pacific during the Cretaceous, when the gap between the Americas was open but species exchange gradually became less during the Tertiary as the Panama Isthmus rose. In the Hawaii group, coral reefs became established only during the Oligocene following the intensifying of the North Pacific ocean-surface gyral circulation (Grigg 1988). Subsequent changes in species composition may be an effect of episodes of extinction and recolonization associated with Quaternary climate changes.

As temperatures and sea levels oscillated during the Quaternary, coral reefs were alternately exposed and drowned. During glacial periods, when sea levels were low, the distribution of coral reefs was much less and in marginal areas of the modern coral seas (like the Hawaiian Islands; Grigg 1988) reefs died out entirely. Owing to cooler temperatures, coral reefs grew at slower rates, and many were comparatively ephemeral. As temperatures increased and sea levels rose at the end of the glacial periods, reefs gradually became reestablished across wider areas of the coral seas. Depending on oceanographic factors, upward-growing coral reefs were either able to 'keep-up' with rising postglacial sea level, or they later were able to 'catch-up', or they had to 'give-up' and thereby forming a drowned reef (Neumann and MacIntyre 1985).

Drowned reefs occur in many parts of the Pacific and Indian Oceans in particular. Most are thought to have failed to keep up with rising sea level during a period of sea-level rise, for reasons associated with climate and sea-level history, palaeolatitude, seawater temperature, and light (Flood 2001). Other 'drowned' coral reefs, particularly those on the flanks of the Hawaiian Ridge, have slipped hundreds of metres downslope.

In many parts of the world, but especially near convergent plate boundaries, coral reefs are found raised above their modern counterparts and, as such, often provide important insights into reef structure and history (Plate 28). Emerged reef staircases on islands like Sumba in Indonesia have been studied in detail (Pirazzoli *et al.* 1991).

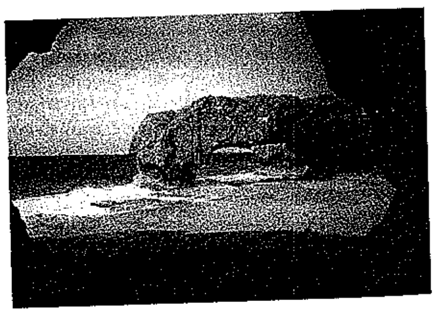

Plate 28 The Talava Arches on Niue Island, central South Pacific. Niue is a fine example of an uplifted atoll, with a well-preserved atoll reef (now 70 m above the modern reef) and lagoon floor. Around the fringes of the emerged atoll reef are a series of emerged fringing reefs. The emerged reef shown here dates from the Last Interglacial period. The modern reef here is rising and is consequently narrow except in embayments as shown here (Photo by Patrick D. Nunn)

In the Pacific and Indian Oceans during the Holocene, keep-up coral reefs grew above their present levels around 4,000 cal. yr BP and have since been exposed as sea level fell. These fossil reefs form the cores of many reef islands (see CAY) in the central Pacific and have been critical factors in their habitability and persistence.

On the basis of their form, coral reefs can generally be either FRINGING REEFS, barrier REEFS or ATOLL reefs. Fringing reefs are juvenile, sometimes ephemeral, and grow outwards from a coast. Barrier and atoll reefs are older, often being composed of reefs of many different ages; reef upgrowth during postglacial periods has been followed by subaerial exposure and erosion during the following glacial period, followed by renewed upgrowth. A good example is that of Midway Island in Hawaii where reef dating back to the mid-Tertiary has been cored (Lincoln and Schlanger 1987). Barrier reefs are separated from a nearby coast by a lagoon while atoll (or ring) reefs enclose a lagoon. The three types were first linked by Darwin (1842) in his Theory of Atoll Development. In this he envisaged that a young volcanic island would develop a fringing reef. As the island subsided so the fringing reef would become transformed into a barrier reef and finally, as the last vestiges of the island were submerged, an atoll reef. Deep drilling of atolls demonstrated the essential correctness of Darwin's model.

Coral reefs in geomorphological research

Since corals are temperature-sensitive organisms, we can learn a lot about palaeoclimates from studying their fossil distributions (see above). We can also use coral reefs as palaeosea-level indicators, both for the Last Interglacial when it is of interest to know whether or not sea level reached 6 m above its present level, as suggested by studies of the emerged reef series on the Huon Peninsula of Papua New Guinea (Chappell 1983). In the Pacific, studies of Holocene emerged reefs have given us much information about the sea-level maximum about 4,500 cal. yr BP (see Nunn and Peltier 2001, for example) and another about 650 cal. yr BP which marked the start of the 'AD 1300 Event' (Nunn 2000). There have also been successful studies of stable isotopes in long-living corals to generate climate data prior to the start of the instrumental record in key regions such as the South Pacific (Quinn *et al.* 1993).

Much research has focused on modelling the relationship between coral reefs and the shorelines which they commonly adjoin, particularly in terms of sediment production, lagoonal dynamics, beach nourishment and shoreline erosion; good studies are those of Munoz-Perez *et al.* (1999) and Hearn and Atkinson (2001). It is clear, for example, that along many tropical coasts, coral reefs are the main producers of the fine-grained sediments which supply nearby beaches and that, should those reefs become degraded, then these beaches can become starved of sediment and destabilized.

Human impacts on coral reefs

It has been realized only comparatively recently how fragile coral-reef ecosystems are, and how much they have been affected by and/or are vulnerable to a variety of human impacts, direct and indirect (Bryant *et al.* 1998). Recently evidence has been presented suggesting that the first human colonizers of some remote Pacific Island groups ~3,000 years ago inadvertently brought with them alien organisms which occupied coral reefs causing reef-surface growth to cease for several hundred years (Nunn 2001). Modern human impacts are more familiar and better understood. These include direct impacts, ranging from the overexploitation of reef organisms (including corals) for sustenance or sale to the dynamiting of reef waters to maximize fish catches, which commonly cause structural reef damage. Indirect impacts are from pollution, including excessive sediment inputs from logging into nearshore areas and chemical pollutants from mineral processing or domestic waste disposal, for example.

Many coral reefs have become degraded as a result of such impacts, manifested as a loss of corals and associated reef organisms, and a reduction in species diversity. Certain more hardy organisms such as sea grasses and various algae (especially *Halimeda*) often cover such degraded reefs. Sometimes reef degradation allows reef predators like the crown-of-thorns starfish (*Acanthaster*) sufficient access to result in an infestation which then exacerbates the process of degradation.

Tourism along tropical coasts often focuses on coral reefs; around 80 per cent of the visitors to the Maldives in 2001 wanted to dive on their reefs. While reef-associated tourism can be sustainable, in many cases it is not because the effects of constructing tourist infrastructure and the effluents which must be disposed of when a large hotel or resort exists in a particular place all reduce the health of the reef ecosystem.

Coral-reef conservation initiatives, including the establishment of marine-protected areas, are often well intentioned but ineffective. Good examples are found in parts of the Caribbean and tropical Pacific Islands where the idea of marine reserves is anathema to people who have been accustomed to free access to reef areas for susbistence purposes (Birkeland 1997).

The future of coral reefs and the implications for geomorphology

Many coral-reef ecosystems have become significantly degraded as a result of human impacts (see above). Many reefs are now being pushed to the brink of extinction because of the additional stress associated with rising sea-surface temperatures (Hoegh-Guldberg 1999). High levels of stress often cause corals to become bleached, the loss of colouration being associated with the ejection of the symbiotic algae that live within coral polyps. Whole reefs can die as a result of bleaching episodes, and there are no instances where a formerly bleached reef has been able to recover its former state. As sea-surface warming continues over the next few decades, so the instances of bleaching resulting from prolonged periods of high temperatures (often associated with El Niño) are likely to increase. The Great Barrier Reef is likely to be experiencing annual bleaching events by 2030.

The implications of regular bleaching for the world's coral reefs are extremely serious, and will have huge implications for many subsistence coastal dwellers in the tropics, who depend daily on reefs for sustenance, and for those countries which depend heavily on revenue generated from reef-associated tourism. The effects for coastal landscapes will involve drastic reductions in the amounts of fine calcareous sediment being generated in reef-lagoon areas, perhaps with many beaches disappearing as a result. This may in turn increase the vulnerability of sandy shorelines to erosion, also an effect of larger waves crossing reefs which are unable to grow upwards in response to projected sea-level rise (Birkeland 1997).

References

Birkeland, C.E. (ed.) (1997) *Life and Death of Coral Reefs*, New York: Chapman and Hall.

Bryant, D., Burke, L. and McManus, J. (1998) *Reefs at Risk: A Map-based Indicator of Threats to the World's Coral Reefs*, Washington, DC: World Resources Institute.

Chappell, J. (1983) A revised sea-level record for the last 300,000 years from Papua New Guinea, *Search* 14, 99–101.

Darwin, C.R. (1842) *Structure and Distribution of Coral Reefs*, London: Smith, Elder.

Flood, P.G. (2001) The 'Darwin Point' of Pacific Ocean atolls and guyots: a reappraisal, *Palaeogeography, Palaeoclimatology, Palaeoecology* 175, 147–152.

Grigg, R.W. (1988) Paleoceanography of coral reefs in the Hawaiian–Emperor chain, *Science* 240, 1,737–1,743.

Grigg, R.W. and Hey, R. (1992) Paleoceanography of the tropical eastern Pacific Ocean, *Science* 255, 172–178.

Hearn, C.J. and Atkinson, M.J. (2001) Effects of sea-level rise on the hydrodynamics of a coral reef lagoon: Kaneohe Bay, Hawaii, in J. Noye and M. Grzechnik (eds) *Sea-Level Changes and their Effects*, 25–47, Singapore: World Scientific Publishing.

Hoegh-Guldberg, O. (1999) Coral bleaching, climate change and the future of the world's coral reefs, *Review of Marine and Freshwater Research* 50, 839–866.

Lincoln, J.M. and Schlanger, S.O. (1987) Miocene sea-level falls related to the geologic history of Midway atoll, *Geology* 15, 454–457.

Munoz-Perez, J.J., Tejedor, L. and Medina, R. (1999) Equilibrium beach profile model for reef-protected beaches, *Journal of Coastal Research* 15, 950–957.

Neumann, A.C. and MacIntyre, I. (1985) Reef response to sea-level rise: keep-up, catch-up or give-up', in *Proceedings of the 5th International Coral Reef Congress* 3, 105–110.

Nunn, P.D. (2000) [illegible]ating sea-level fall around AD 1220–1510 (730–4[illegible]al. yr BP) in the Pacific Islands: implications for env[illegible]mental change and cultural transformation, *New Zealand Geographer* 56, 46–54.

——(2001) Ecological crises or marginal disruptions: the effects of the first humans on Pacific Islands, *New Zealand Geographer* 57, 11–20.

Nunn, P.D. and Peltier, W.R. (2001) Far-field test of the ICE-4G (VM2) model of global isostatic response to deglaciation: empirical and theoretical Holocene sea-level reconstructions for the Fiji Islands, Southwest Pacific, *Quaternary Research* 55, 203–214.

Pirazzoli, P.A., Radtke, U., Hantoro, W.S., Jouannic, C., Hoang, C.T., Causse, C. and Best, M.B. (1991) Quaternary raised coral-reef terraces on Sumba Island, Indonesia, *Science* 252, 1,834–1,836.

Quinn, T.M., Taylor, F.W. and Crowley, T.J. (1993) A 173 stable isotope record from a tropical South Pacific coral, *Quaternary Science Reviews* 12, 407–418.

SEE ALSO: atoll; fringing reef; reef

PATRICK D. NUNN

CORNICHE

Corniches are narrow organic ledges, 0.5 to 2 m in width, growing on steep rock surfaces at about mean sea level. The best examples are on limestones where notches (see NOTCH, COASTAL) develop in the spray zone. Corniches in the north-western Mediterranean consist of algae, particularly the calcareous alga *Tenarea tortuosa*, although Serpulid (see SERPULID REEF) worms or Vermetid (see VERMETID REEF AND BOILER) gastropod tubes can play a similar role. Although the interiors are generally quite hard, corniches cannot resist very strong waves and they are best developed in inlets on exposed coasts.

Reference

Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Oxford University Press.

ALAN TRENHAILE

CORROSION

Corrosion is synonymous with solutional erosion, the erosion of material by chemical activity. The majority of studies of corrosion have been undertaken on carbonates and these are the primary focus of this entry. However, similar considerations apply to evaporites and the estimation of gypsum corrosion rates is particularly problematic because of the more rapid solution and the consequently greater spatial and temporal variability of dissolution.

Corrosion rates are commonly expressed as mm/1000a, implying that all corrosion contributes to surface lowering and that environmental conditions have remained broadly the same for millennia. The former is incorrect, particularly in karst, while the latter is also highly questionable. The preferred unit is $m^3\ km^{-2}\ a^{-1}$ and $1\ mm\,ka^{-1}$ is equivalent to $1\ m^3 km^{-2} a^{-1}$. Where surface lowering is measured directly then units of $\mu m\ a^{-1}$ are appropriate.

Limestone corrosion rates may be estimated from knowledge of dissolution kinetics, runoff, carbon dioxide and temperature, but there remains a need for field measurements to provide actual values of regional denudation; to compare rates in contrasting environments and by different processes; to understand landform evolution; and to understand how processes operate in a complex natural environment as opposed to the laboratory. When evaluating results from past studies it is important to understand what was actually measured and how the corrosion rates were calculated. Most field measurements of corrosion in carbonate karst were based on spot samples, with denudation being estimated from the Corbel formula. This suffers several problems, the three most important being: the carbonate concentration is frequently the average of a few spot measurements, with the implicit assumption of a linear relationship between carbonate hardness and discharge; carbonates present in solution are assumed to only come from karst denudation; and measurements are usually made at only one point, commonly the output of a drainage basin, with the implicit assumption that this is representative of conditions upstream.

Where water samples have been collected over a range of flow conditions it is apparent that the relationship between dissolved load and discharge is usually non-linear and particularly in small drainage basins may be complicated by hysteresis effects (usually higher concentrations per unit discharge on the rising limb). It is virtually impossible to correct for hysteresis, but by collecting samples over a range of discharges it is usually possible to construct a reliable discharge-concentration or discharge-load rating curve. This can be applied to the discharge curve and the results summed to obtain the total annual solute load. Greater accuracy may be obtained using a logging conductivity meter, developing a conductivity-concentration rating curve, and using this to predict the concentration at each measured discharge.

Having computed the total solute load (TSL) at a point it is important to realize that this is made up of corrosion of karst rocks by both autogenic (CKAu) and allogenic waters (CKAl), less any deposition of previously dissolved material (D), together with corrosion of non-karst rocks by allogenic waters (CNK), solute accessions in rainfall and snowfall (AC), and any anthropogenic inputs such as fertilizers (AN). The gross karst solution is then (CKAu + CKAl) whereas the net karst solution is (CKAu + CKAl − D). Where precipitation of previously dissolved carbonates is minimal then gross and net solution will be similar, but elsewhere failure to account for deposition may result in a significant underestimate of gross denudation, which is the real measure of relief transformation. In contrast, failure to take into account the solution of non-karst rocks and solute accessions in precipitation will result in an overestimate of karst corrosion. Error in estimating corrosion rates can arise from many sources and even in a careful study using hydrochemical budgeting and taking into account non-denudational components potential errors of around 25 per cent are possible.

Corrosion rates for whole drainage basins derived by sampling of water at the basin outlet are unlikely to be representative of any specific location within the basin. This information may best be obtained by an extension of the hydrochemical budgeting method discussed above. Water samples are collected from the full range of sites – bare limestone surfaces, the soil zone, the subcutaneous zone, the main body of bedrock, and cave streams in both vadose and phreatic zones. These, together with estimates of the proportion of water following the various pathways through the system, permit the breaking down of the overall corrosion budget. Those few studies that have been made show that a high proportion of corrosion (50–85 per cent) occurs within several metres of the surface in the soil (if present) and subcutaneous zone (uppermost bedrock). Caves account for very little of the erosion when averaged over the whole basin.

The principal drawback of the hydrochemical approach is that it requires frequent, ideally continuous, measurement of discharge and sufficient samples to establish the pattern and extent of variations in solute concentrations. As this is not always possible alternative methods that integrate erosion over a longer time period have been derived. The two most commonly used are the micro-erosion meter and rock tablets. In contrast to the hydrochemical method these are highly site-specific and may only be used to assess corrosion rates on bare limestone surfaces, in the soil zone, at the soil–bedrock interface, and in cave streams. Tablets have been found to give estimates two orders of magnitude less than those calculated using hydrochemical data. The most likely explanation is that natural rock surfaces come into contact with larger volumes of water than do isolated rock tablets, simply because of their greater lateral flow component. Thus, the two methods measure fundamentally different phenomena and the hydrochemical method provides the only reliable means of estimating corrosion rates on limestone surfaces. Different problems arise if tablets are placed in cave streams as they will project above the natural surface and as a consequence are likely to erode more rapidly. They are also likely to suffer from abrasion as well as corrosion, although this can be exploited by placing the tablets in nylon cages with differing mesh sizes and comparing the erosional losses suffered.

Further reading

Dreybrodt, W. (1988) *Processes in Karst Systems: Physics, Chemistry and Geology*, Berlin and New York: Springer.

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

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

SEE ALSO: dissolution

JOHN GUNN

COSMOGENIC DATING

Cosmogenic dating is a group of related techniques for estimating landform ages and erosion rates. It is based upon the generation of rare isotopes within minerals by cosmic rays. Primary cosmic rays composed largely of highly energetic protons interact with gases in the Earth's atmosphere to produce showers of secondary subatomic particles, mostly neutrons and muons. These secondary cosmic rays induce nuclear reactions within the Earth's terrestrial surface, producing cosmogenic nuclides. The length of surface exposure, or alternatively, the rate of surface erosion, is computed from the concentration of cosmogenic nuclides in a landform.

Six cosmogenic nuclides have found widespread application in geomorphology (Table 9). These are stable isotopes of the noble gases helium and neon (^{3}He and ^{21}Ne), and radioactive isotopes of beryllium, carbon, aluminum, and chlorine (^{10}Be, ^{14}C, ^{26}Al, and ^{36}Cl). The nuclides ^{10}Be, ^{14}C and ^{36}Cl are also produced within the atmosphere by cosmic rays. The best known example of atmospheric production is ^{14}C which forms the basis for radiocarbon dating (see DATING METHODS). To avoid confusion, nuclides generated within mineral lattices in the Earth's solid surface are termed *in situ*-produced terrestrial cosmogenic nuclides (TCN).

Most TCN production is from neutron spallation (Lal 1991). TCN spallation occurs when a

Table 9 Properties of *in situ*-produced terrestrial cosmogenic nuclides (TCN)

Nuclide	Mean lifetime (yrs)	Host mineral
^{3}He	Stable	Olivine, clinopyroxene
^{21}Ne	Stable	Olivine, clinopyroxene, quartz
^{10}Be	2.2 Myr	Quartz
^{14}C	0.82 kyr	Quartz, calcite
^{26}Al	1.0 Myr	Quartz
^{36}Cl	430 kyr	Calcite, dolomite, whole rocks

secondary neutron with energy >10 MeV collides with a target nucleus in a mineral lattice, breaking protons, neutrons or clusters of these particles from the nucleus. Spallation products always consist of an isotope of lower atomic number than the target. As neutrons do not penetrate deeply in rocks, most neutron spallation occurs within about one metre of the surface. Thermal neutrons (energy ~0.025 eV) are absorbed by some target nuclei, causing radioactive decay and production of a cosmogenic isotope. Thermal neutron production is important for cosmogenic ^{36}Cl. Muons also create cosmogenic nuclides but at rates much lower than neutron spallation. Muons penetrate far more deeply than neutrons, creating measurable quantities of cosmogenic nuclides at depths of over 20 metres (Granger and Muzikar 2001).

The production rate of cosmogenic nuclides by all reaction mechanisms is low, ranging (at sea level and latitudes $>60°$) from about 5 to 6 atoms $g^{-1}\,a^{-1}$ for ^{10}Be to about 120 atoms $g^{-1}\,a^{-1}$ for ^{3}He. Cosmic rays are attenuated by the atmosphere and the geomagnetic field; consequently production rates vary significantly with altitude and latitude. For this reason, TCN production rates are always quoted for sea level and high latitude, and scaled to the altitude and latitude of study sites using empirical functions (Lal 1991; Stone 2000; Dunai 2000). Production rates must be precisely known for reliable TCN results. This is a difficult task because both atmospheric shielding and geomagnetic field intensity vary with time. Calibration sites are used to determine production rates. At a calibration site, TCN concentrations are measured in independently dated geomorphic surfaces with near-zero erosion rates such as lava flows, glacially eroded bedrock or large landslides.

Applications of TCN fall into two main categories: surface exposure dating and measurement of erosion rates. Both applications yield model results with accuracy highly dependent on the validity of simplifying assumptions. In exposure dating, the first requirement is that the geomorphic surface must have formed over a short time period. Examples of such surfaces include fault scarps (see FAULT AND FAULT SCARP), LAVA LANDFORMS, LANDSLIDES and ERRATIC boulders. Surfaces forming incrementally over long periods of time have cosmogenic nuclide concentrations best interpreted in terms of erosion rates. The second requirement is that the geomorphic surface be free of TCN at the time of surface formation. Remnant TCN from past periods of surface exposure is termed nuclide inheritance. Lava flows and large glacial erratics generally have little or no nuclide inheritance. The final requirement for accurate exposure dating is that the primary geomorphic surface form must be preserved over the period of exposure. Erosion rates must either be known, or be assumed to be zero. Surface exposure dating therefore requires careful analysis of landscapes and sampling of surfaces experiencing very low rates of erosion. The requirement of near-zero erosion limits the age range of TCN exposure dating. In most geomorphic environments, reliable exposure ages generally range from about 5,000 years to less than 100,000 years. Younger ages are limited by detection limits for measuring TCN while older surfaces are generally destroyed by erosion or buried by sediments. The polar deserts of east Antarctica are a major exception, with exposure ages of over 5 million years. The precision of exposure ages and erosion rates, as estimated by analytical errors, varies with isotope and application but generally ranges between ±3 per cent to 15 per cent.

In TCN erosion rate studies, an assumption of equilibrium between TCN production and loss by erosion and radioactive decay is made (Bierman and Steig 1996; Granger *et al.* 1996). Under these circumstances, exposure time is not important and TCN concentrations vary inversely with erosion rates. For example, steep hillslopes with high erosion rates have low TCN concentrations because of the short residence time of target minerals within the zone of production. Averaging time is the time necessary to achieve equilibrium conditions. The lower the erosion rate, the longer the averaging time. TCN averaging times for erosion rates typical of temperate climates range from about 100,000 years to about 5,000 years. Averaging erosion over such timescales makes the TCN method very useful for investigating links between climate and tectonics, and for establishing baseline erosion rates unrelated to human activities. Two types of sampling are applied in TCN erosion rate studies. Bedrock samples give information about minimum rates of landscape lowering and the influence of lithology on erosion rates. Alluvial samples average erosion rates for the contributing catchment and therefore are easiest to compare with traditional methods of measuring short-term erosion rates such as suspended sediment studies.

TCN vary greatly in terms of ease and cost of measurement, sample preparation and host minerals. ^{3}He and ^{21}Ne are measured in olivine and clinopyroxene using noble gas mass spectrometry techniques similar to those employed for $^{40}Ar/^{39}Ar$ studies. They are primarily used for dating mafic volcanic rocks, and for studies of long-term landscape evolution in Antarctica where extremely low rates of erosion require the use of a stable nuclide (Summerfield *et al.* 1999). The most used TCN are ^{10}Be and ^{26}Al. These nuclides are popular because the host mineral (quartz) is present in the majority of geologic settings, production reactions are relatively simple and well understood, and both nuclides can be measured in the same sample. Since the mean lives of ^{10}Be and ^{26}Al differ significantly (Table 9), measurement of the nuclides in the same sample can constrain both erosion rate and exposure time as well as indicate periods of burial (Lal 1991; Bierman *et al.* 1999). It is also possible to use this nuclide pair for dating the burial of sediments (Granger and Muzikar 2001). Measurement of ^{10}Be and ^{26}Al is by accelerator mass spectrometry (AMS). Sample preparation requires preparation of high purity quartz separates and removal of atmospheric ^{10}Be with hydrofluoric acid etching. ^{36}Cl is also widely applied to geomorphic problems, particularly in carbonate and volcanic landscapes where ^{10}Be cannot be used. Production rates for ^{36}Cl are less well established than for other TCN because of more complex production reactions. ^{36}Cl is produced by neutron spallation on K and Ca, and by thermal neutron capture on ^{35}Cl. Production rates vary with rock composition, and major and trace element data are needed to compute rates. AMS is also used to measure ^{36}Cl concentrations. *In situ*-produced ^{14}C has not been widely applied in geomorphology because of problems separating atmospheric contamination. However, applications with this nuclide are likely to increase. The mean life of ^{14}C is much shorter than any other TCN, therefore, when measured in conjunction with ^{10}Be and ^{26}Al, it can be used to establish production rates, estimate erosion corrections, and detect periods of burial by sediment or ice.

References

Bierman, P.R. and Steig, E.J. (1996) Estimating rates of denudation using cosmogenic isotope abundances in sediment, *Earth Surface Processes and Landforms* 21, 125–139.

Bierman, P.R., Marsella, K.A., Paterson, C., Davis, P.T. and Caffee, M. (1999) Mid-Pleistocene cosmogenic minimum age limits for pre-Wisconsin glacial surfaces in southwestern Minnesota and southern Baffin Island: a multiple nuclide approach, *Geomorphology* 27, 25–40.

Dunai, T.J. (2000) Scaling factors for production rates of *in situ* produced cosmogenic nuclides: a critical reevaluation, *Earth and Planetary Science Letters* 176, 157–169.

Granger, D.E. and Muzikar, P.F. (2001) Dating sediment burial with *in situ*-produced cosmogenic nuclides: theory, techniques, and limitations, *Earth and Planetary Science Letters* 188, 269–281.

Granger, D.E., Kirchner, J.W. and Finkel, R. (1996) Spatially averaged long-term erosion notes measured from *in-situ*-produced cosmogenic nuclides in alluvial sediments, *Journal of Geology* 104(3), 249–257.

Lal, D. (1991) Cosmic ray labeling of erosion surfaces: *in situ* nuclide production rates and erosion rates, *Earth and Planetary Science Letters* 104, 424–439.

Stone, J.O. (2000) Air pressure and cosmogenic isotope production, *Journal of Geophysical Research* 105, 23,753–23,759.

Summerfield, M.A., Stuart, F.M., Cockburn, H.A.P., Sugden, D.E., Denton, G.H., Dunai, T. and Marchant, D.R. (1999) Long-term rates of denudation in the Dry Valleys, Transantarctic Mountains, southern Victoria Land, Antarctica, based on *in-situ* produced cosmogenic ^{21}Ne, *Geomorphology* 27, 113–130.

Further reading

Bierman, P.R. (1994) Using *in situ* produced cosmogenic isotopes to estimate rates of landscape evolution: a review from the geomorphic perspective, *Journal of Geophysical Research* 99, 13,885–13,896.

Cerling, T.E. and Craig, H. (1994) Geomorphology and *in-situ* cosmogenic isotopes, *Annual Reviews of Earth and Planetary Sciences* 22, 273–317.

Gosse, J.C. and Phillips, F.M. (2001) Terrestrial *in situ* cosmogenic nuclides: theory and application, *Quaternary Science Reviews* 20, 1,475–1,560.

WILLIAM M. PHILLIPS

COULEE

In western North America, coulee (French *couler*: to flow) is a common term used to describe a dry valley, canyon, gulch or wash. Most coulees were formed rapidly in late glacial times by large discharges of melt water, particularly with the emptying of proglacial lakes (Bretz 1969). Selby (1985: 458) adopts this as a specific origin. Coulees may have ponded water bodies, intermittent or underfit streams. Parallel sets of coulees in southern Alberta, Canada, may have been aligned by regional joint patterns (Babcock 1974), or possibly formed in postglacial time through some imperfectly understood process controlled by prevailing wind direction (Beaty 1975). Less commonly, the term coulee is used to describe a short lobe of viscous lava on the flanks of a volcano and a lobe of debris moved by gelifluction.

References

Babcock, E.A. (1974) Photolineaments and regional joints: lineament density and terrain parameters, south-central Alberta, *Bulletin of Canadian Petroleum Geology* 22, 89–105.

Beaty, C.B. (1975) Coulee alignment and the wind in southern Alberta, Canada, *Geological Society of America Bulletin* 86, 119–128.

Bretz, J.H. (1969) The Lake Missoula floods and the channelled scablands, *Journal of Geology* 77, 505–543.

Selby M.J. (1985) *Earth's Changing Surface*, Oxford: Clarendon Press.

ROBERT J. ROGERSON

COVERSAND

Originally a Dutch term applied to aeolian SANDSHEET deposits overlying older sediments. Its generic nature has led it to be applied to a range of deposits. However, the common denominator has been its application to sandsheet deposits of cold-climate (see PERIGLACIAL GEOMORPHOLOGY) aeolian origin. The latter is proven by the occurrence in coversands of frost cracks, involutions and ice wedge casts (see ICE WEDGE AND RELATED STRUCTURES), as well as from pollen and beetle evidence obtained from intercalated organic deposits. The aeolian origin of coversands has been determined on the basis of their concordance with dune forms (see DUNE, AEOLIAN), occurrence with VENTIFACTS and/or on the basis of particle characteristics (mineralogy, sorting, rounding, surface matting and textures). Whilst predominantly aeolian derived, coversands can incorporate components of sand derived from other processes, e.g. niveo- and/or fluvio-aeolian (see NIVEO-AEOLIAN ACTIVITY).

Coversands in northern Europe (Schwan 1988) and mid-continental north America although relict, are widespread, extending over 10,000s of km^2 as nearly spatially continuous sheets with flat to undulating relief (less than 5 m) and a notable paucity of dunes (Koster 1988). This differentiates coversand from more recent sand deposits which tend to have been formed into dunes, e.g. Drift sands in The Netherlands (Koster *et al.* 1993). The coversand is typically of a uniform thickness of up to several metres; only in valleys, depressions or against topographical barriers is it thicker. The coversands also tend to be (sub)horizontally stratified, composed of thin beds, setting them apart from the high angle bedding of coastal dune (see DUNE, COASTAL) sands and the cross bedding, troughs and ripples of riverine sands.

Detailed examination of coversand stratification in Europe has led to classification of coversands into two types which are in turn subdivided into two: Older coversands I and II and Younger coversands I and II. The Older coversand is characterized by an alternation of well-sorted parallel-laminated beds of greyish loam/fine sand and yellowish fine/medium sand. The Older coversand I has evidence of more cryogenic deformation and frost wedge casts, especially in its upper layers, than Older coversand II and the two facies are commonly separated by a disconformity, e.g. the Beuningen pebble. The Younger coversand is typically a unimodal, well-sorted, parallel-laminated medium sand with a large sand component derived from local sources. The sand is rarely buried or cross-bedded, has a low relief and has no evidence of ice wedge formation in it. The primary differentiation between the two Younger coversand facies is on sedimentary structures which indicate that the Younger coversand II was deposited under drier conditions.

Fragmentary evidence indicates coversands have been deposited during several Pleistocene glacials and are not unique just to the last glacial cycle. The northerly limit of the relict but extensive European coversands found in Britain, The Netherlands, Germany, Denmark, Poland and the Baltic states is broadly coincident with the maximal position of the Late-Weichselian (Devensian) ice sheet. In general, the last era of north-west European coversand activity started after the last interglacial, increasing in intensity throughout the Weichselian period. Two main phases of coversand deposition have been reported: one around 18,000–15,000 years ago (Older Coversand II) and another more intense period between 14,000–11,000 years ago (Younger Coversand) (Koster 1988; Bateman 1998; Singhvi *et al.* 2001). Older coversand I appears to have been dominantly deposited separate to, but contemporary with, the widespread LOESS deposits of north-western and eastern Europe which were mostly deposited just prior to the last glacial maximum and appear to have stabilized by approximately 13,000 years ago (Singhvi *et al.* 2001). However, evidence also suggests localized environmental conditions blurred these discrete aeolian phases with Older coversand type facies still being deposited in places during the so-called Younger coversand phase (Kolstrup *et al.* 1990; Kasse 1997).

Formation and preservation of the Late-Weichselian coversands is thought to have been aided by enhanced sand sources as a result of glaciation, sparse vegetation, low relief and low sand supply due to periodically wet, frozen or cemented depositional surfaces (Kasse 1997). Use of the orientation of dune morphology, bedding inclination and unit thickness has enabled the reconstruction of palaeo-wind directions. The Older coversands type was deposited by predominantly north-westerly to westerly winds and the Younger coversands deposited by more westerly to south-westerly winds. Such information has been used to inform palaeoclimatic models for north-western and central Europe (e.g. Isarin *et al.* 1998).

References

Bateman, M.D. (1998) The origin and age of coversand in north Lincolnshire, UK, *Permafrost and Periglacial Processes* 9, 313–325.

Isarin, R.F.B., Renssen, H. and Vandenberghe, J. (1998) The impact of the North Atlantic Ocean on the Younger Dryas climate in northwestern and central Europe, *Journal of Quaternary Science* 13, 447–453.

Kasse, C. (1997) Cold-Climate aeolian sand-sheet formation in North-Western Europe (*c.*14–12.4 ka); a response to permafrost degradation and increased aridity, *Permafrost and Periglacial Processes* 8, 295–311.

Kolstrup, E., Grun, R., Mejdahl, V., Packman, S.C. and Wintle, A.G. (1990) Stratigraphy and thermoluminescence dating of Late Glacial cover sand in Denmark, *Journal of Quaternary Science* 5, 207–224.

Koster, E.A. (1988) Ancient and Modern cold-climate aeolian sand deposition: a review, *Journal of Quaternary Science* 3, 69–83.

Koster, E.A., Castel, I.I.Y. and Nap, R.L. (1993) Genesis and sedimentary structures of Late Holocene aeolian drift sands in northwest Europe, in K. Pye (ed.) *The Dynamics and Environmental Context of Aeolian Sedimentary Systems*, Geological Society Special Publication 72, 247–267.

Schwan, J. (1988) The structure and genesis of Weichselian to early Holocene aeolian sand sheets in W. Europe, *Sedimentary Geology* 55, 197–232.

Singhvi, A.K., Bluszcz, A., Bateman, M.D. and Someshwar Rao, M. (2001) Luminescence dating of loess-palaeosol sequences and coversands: methodological aspects and palaeoclimatic implications, *Earth-Science Reviews* 54, 193–211.

MARK D. BATEMAN

CRATER

Craters are bowl-shaped, approximately circular depressions that typically form by high-energy impact or explosive activity. There is a fundamental geomorphological problem in distinguishing crater origins by volcanic versus impact processes. The latter involve collision with a planetary surface by meteors, comets and asteroids. It is also possible that a crater can form by the explosion, just above the ground, of a meteor or comet, known in this context as a *bolide*. Volcanic craters generally form at the summits of volcanic cones and result from explosive eruptions or the accumulation of *pyroclastic* material in a rim around a volcanic vent. Of course, human activity can produce explosion craters, perhaps the most spectacular of which resulted from nuclear testing. Interestingly, it was the well-funded study of physics for the latter that ultimately led to considerable advancement in understanding the natural process of impact cratering (Roddy *et al.* 1977).

One of the great controversies in planetary geomorphology concerned the origin of craters on the moon. G.K. Gilbert (1893) used geomorphological reasoning to argue that the moon's craters had an impact origin. Until the 1930s, however, most astronomers thought that the circularity of the moon's craters required an origin by volcanic processes. Objects striking the moon, it was thought, would include many oblique impacts, and these would not be circular. Only later in the twentieth century did the physics of the cratering process come to be well understood enough to show that most oblique impacts produced circular, rather than elliptical, craters. Nevertheless, some astronomers continued to argue for the volcanic origin until the Apollo missions of the 1970s returned incontrovertible proof of the impact origin for nearly all lunar craters.

Volcanic craters

Craters can be a variety of depressions associated with volcanic or pseudovolcanic activity, including mud volcanoes, mound springs, hot springs, and even pingos. The geomorphology of truly volcanic craters was reviewed by Fairbridge (1968), who considered the large complex collapse and explosion structures known as *calderas* separately from other volcanic craters. Magmas rich in silica tend to produce highly explosive activity in which the volcanic materials become fragmented into *pyroclastic* rock. Domes of silica-rich volcanic rock, including obsidian, commonly fills pre-existing pyroclastic craters. Explosive activity for basaltic magmas produces spatter cones with craters over rift zones, and a variety of pit craters. One of the most famous of these is Halemaumau, a pit crater on the floor of Kilauea Caldera, on Earth's most continuously active volcano at the southern end of the island of Hawaii. There are also many volcanic craters on other planets, including spectacular calderas on the volcanoes of Venus, Mars and Io (the highly volcanically active satellite of Jupiter).

A special type of crater, known as a *maar*, derives its name from the Rhineland dialect of German. The term was originally applied to volcanic explosion craters near Eifel, Germany. Maar craters may be associated with *diatremes*, which are breccia-filled volcanic pipes that form by gas explosions. They also occur within fields of monogenetic volcanic cones, which develop during single eruptive phases. Maar craters usually have a ring of erupted pyroclastic material, and lakes often occur on their floors. They generally form by the interaction of rising lavas with near-surface ground water.

Another interesting crater form is known as a pseudocrater, or rootless cone. These were first recognized in Iceland, where basaltic lava flows advanced over substrates that were rich in water or ice. The interaction of the lava and water produced pyroclastic explosions that formed the craters. The advancing flow may then separate the crater or cone from its source zone. Such features range from a few tens of metres to hundreds of metres in diameter.

Impact crater morphology

It is one of the major discoveries of recent planetary exploration that the surfaces of rocky objects in the solar system are almost all marked by numerous impact craters. These occur over an immense range of size scales. The smallest are microcraters or pits, which form from the impact of micrometeorites or high-velocity cosmic dust grains on exposed rocks. These only form on bodies that lack atmospheres, which would induce the very small projectiles to burn up before impact. Simple craters are larger, bowl-shaped depressions that form on land surfaces. They range up to several kilometres across, and typically have diameters across their rims that are about five times their depths from rim top to crater floor. Simple craters are familiar to many geomorphologists because they were much in evidence during the Apollo landings of humans on Earth's moon. On Earth one of the most famous simple craters is the 1-km diameter Barringer Crater in northern Arizona, also known as Meteor Crater. It is interesting that a major controversy occurred in regard to its origin, with Gilbert (1896) eventually concluding that it had a volcanic origin, despite making a strong argument for impact as well.

Most of the larger craters visible on planetary and satellite surfaces are complex craters. These have rims marked by terraces along their inner margins. Their floors are broad and flat, and there is often a central peak. Such craters are generally from a few tens to a few hundred kilometres in diameter, and they are well known from observations of the moon (Figure 28). Because of their flat floors and very high ratios of width to depth, these features are usually described as impact basins, rather than craters. Much larger impact structures are also known, and many of these are multi-ring basins. They have multiple concentric rings, each consisting of rugged hilly terrain. The floors of these exceptionally large craters are commonly flooded by lava. They can have diameters of up to two thousand kilometres or more.

Recent work has shown that many of the projectiles generating impact craters arrive in groups, rather than as single projectiles. One of the most spectacular examples of this phenomenon was the comet Shoemaker-Levy 9, which broke into fragments as it collided with Jupiter in July 1994. Asteroids also may break up when they interact with a planet's atmosphere. Among the 150 or so

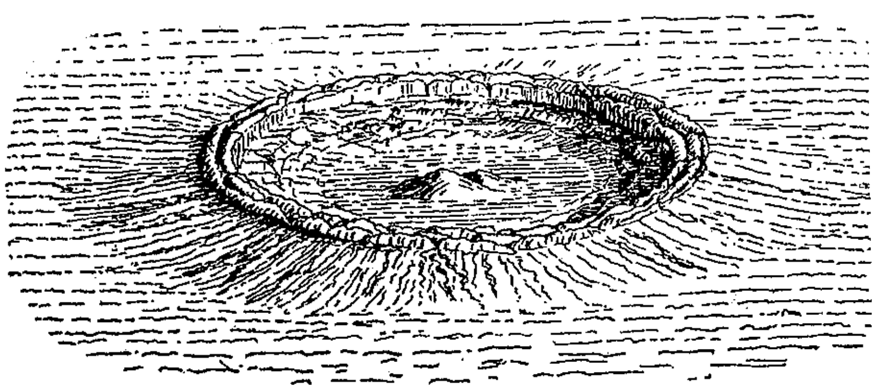

Figure 28 Sketch of a complex lunar crater made by Grove Karl Gilbert (1893: 243)

Plate 29 Henbury impact craters in central Australia. These structures formed when a group of meteors struck a pediment surface less than about 5,000 years ago (Milton 1968). The largest of the craters, part of a tight group of four, is about 150 m in diameter and about 10 to 15 m deep. Note the capture of drainage by the craters

Earth impact sites are many that include multiple craters (Plate 29). The Kaali impacts, which struck Estonia about 2,400 to 2,800 years ago, consist of nine craters, the largest of which is 110 m wide and 20 m deep.

Impact crater processes

Meteors and comets arrive at velocities of many metres per second, causing an immense transfer of energy in an exceedingly short period of time as they strike the surface of a planet. The actual cratering process is surprisingly orderly, and very well known from both theoretical and experimental work (Melosh 1989). The initial phase is contact and compression, which lasts only a few times longer than the time it takes for the projectile to traverse its own diameter. This produces prominent very high-velocity jets of highly shocked material that shoot upward from the margins of the deforming projectile. A zone of phenomenally high pressure is produced at the front of the projectile, as it is deformed by contact with the target material. In the inner solar system the target is usually rock, but in the outer solar system the satellites of Jupiter, Saturn, Uranus and Neptune are commonly icy. The ices are so cold that they generally behave like rock.

Contact and compression is followed by an ejection or excavation phase. The projectile is melted or evaporated by a shock wave propagating into it, while another shock wave propagates into the target. The shock wave is followed by rarefaction waves that decompress the material and generate excavation flows that open up a transient crater. This excavation process may last several minutes, depending on the energy level of the original impact. Material ejected from the crater will then comprise an outwardly expanding ejecta curtain, which has the form of an inverted cone, centred on the impact site. Material deposited from this curtain will comprise an ejecta blanket that covers the terrain out to about two crater radii from the rim. Additional large ejecta blocks may create additional impacts, or secondary craters. These have distinctive morphologies because of the slower projectile velocities, highly oblique paths and radial structure in relation to their source craters.

At the end of the excavation stage the transient crater will often experience collapse and modification. For the larger complex craters this produces terraces and central peaks. The terraces develop by slumping of the crater rim after all material has been excavated. The central peaks represent uplift of the floor material beneath the transient crater cavity. A peak ring may form as the central peak grows and collapses.

Cratered landscapes

Cratered landscapes dominate on the surfaces of rocky objects in the solar system. This is mainly because most of those surfaces are very old. In general, the density of impact craters on a surface corresponds approximately to the age of that surface. However, this relationship holds on very long timescales. Moreover, it is not linear. During the early part of solar system history, the impacting rate was extremely high. From the final accretion of planets and many satellites, about 4.5 billion years ago, until about 3.9 billion years ago for the moon, and perhaps a few hundred million years later for Mars, there was a period of intense heavy bombardment. This produced overlapping craters with sizes up to the scale of the multi-ring basins. The scaling is very regular with many more craters of smaller sizes than of larger. After the heavy bombardment, which was caused by many objects left over from solar system formation, the impacting rates declined by more than an order of magnitude. On the moon these timescales of cratering have been confirmed by radiometric dates on rocks returned to Earth by the Apollo missions. The lunar highlands correspond to the heavy bombardment phase, and much lower crater densities mark the younger volcanic plains of the lunar mare, which occur on the floors of very large impact basins. On Mars there is a similar dichotomy between old, heavily cratered highlands and younger, lightly cratered plains. Unlike the moon, however, many Martian craters are highly degraded by erosion, including the action of fluvial, periglacial and glacial processes.

References

Fairbridge, R.W. (1968) Crater, in R.W. Fairbridge (ed.) *Encyclopedia of Geomorphology* 207–218, New York: Reinhold.

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

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

Melosh, H.J. (1989) *Impact Cratering: A Geologic Process*, Oxford: Oxford University Press.

Milton, D.J. (1968) Structural geology of the Henbury meteorite craters, Northern Territory, Australia, *US Geological Survey Professional Paper* 499-C, 1–17.

Roddy, R.J., Pepin, R.O. and Merrill, R.B. (eds) (1977) *Impact and Explosion Cratering*, New York: Pergamon.

SEE ALSO: astrobleme; caldera; extraterrestrial geomorphology; volcano

VICTOR R. BAKER

CRATON

The central core of extensive, stable continental crust in present-day continents that has achieved tectonic stability. All cratons are older than 570 million years, dating from the Precambrian period. Cratons have essentially rigid foundations, composed of predominantly granite and metamorphosed rocks that have been deposited on pre-existing older basement rocks. They are generally low-lying with little relief, as a result of erosion. Cratons have only been affected by EPEIROGENY and are devoid of orogenic features and recent volcanic activity.

The term craton is derived from the Greek word 'kraton' meaning shield and should only be applied to continents and not to oceans, according to the theory of plate tectonics. Cratons are added to by the process of cratonization, an important mechanism for continental growth. Sediments accumulate within thick linear troughs on the cratonic margins. Here, the material is eventually deformed and partially melted onto the existing craton. Early Achaean cratons were smaller and greater in number, yet through the process of cratonization throughout Phanerozoic time cratons became larger and fewer in number as they were fused together.

The area of a craton that becomes exposed is termed a SHIELD. Shields are composed of ancient crystalline basement rocks, and represent the core of the craton. The Canadian Shield is an example;

it is composed of granite and metamorphic rocks (e.g. gneisses), alongside heavily deformed metamorphosed sedimentary (e.g. quartzites) and other volcanic rocks. The term shield is also sometimes employed as a synonym for craton.

The shield is unconformably overlapped at its margins by thin sedimentary units, termed platforms. Platforms are typically *c.*1 km thick and derived from the Palaeozoic and Mesozoic periods, predominantly composed of shallow marine sandstones, limestones and shales.

Since cratons are tectonically stable, sediments tend to spread out widely into any areas of relatively low-lying ground, such as the intra-cratonic basins. These are typically shallow (though can range up to 3,000 m), bowl-shaped, and are characterized by very slow subsidence. Basin sediments thicken regularly towards the centre, yet their fill is discontinuous. As such, the stratigraphy reflects major transgressions across the entire craton, punctuated by periods of stability. Many of them develop as shallow 'sag' lakes, such as Lake Chad in North Africa. The cause of intra-cratonic basins remains contentious.

Further reading

Condie, K.C. (1997) *Plate Tectonics and Crustal Evolution*, 4th edition, Oxford: Butterworth Heinmann.

STEVE WARD

CROSS PROFILE, VALLEY

In most introductory physical geography and geology textbooks a distinction is made between V-shaped valley cross profiles, described as characteristic of a system dominated by active fluvial erosion, and U-shaped valley cross profiles, described as characteristic of a system dominated by GLACIAL EROSION. This process-oriented distinction gained wide popularity as a component of classic Davisian landscape classification, particularly in the first half of the twentieth century, and continues to be used in more modern landscape interpretation and analysis. Morphometric analyses have been used to show that glaciated valley cross-section profiles can be approximated by the mathematical equivalent of the letter U, a parabolic equation, whereas fluvial valley side slopes are more nearly linear. In addition, the amount of rock removal required to convert a V-shaped cross-profile geometry to a U-shape has been used as a measure of the glacial erosion component of valley development, and the extent of valley development towards a particular form has been used as a measure of the degree of valley modification by fluvial or glacial processes. However, valley cross profiles include a much wider variety of forms than the two-fold division into V- and U-shaped suggests, and cross-profile forms are not only controlled by glacial and fluvial erosion, but also by patterns of hillslope erosion and deposition, and by patterns of rock resistance to erosion.

Typical explanations for the development of V-shaped valleys in areas with active fluvial erosion include several components: (1) that river erosion is dominantly vertical; (2) that the river is capable of transporting all of the material supplied to it by hillslope processes; and (3) that valley side slopes are steepened to a critical angle for hillslope transport or failure. This ideal set of conditions results in uniform valley side slope angles either side of a central river that is eroding vertically into the landscape with little or no floodplain, i.e. a V-shaped cross profile. However, if the river is not capable of transporting all the material supplied to it by hillslope processes, if there are significant lithological variations along the slope profile, or if different hillslope processes dominate in separate parts of the slope profile, then more complex hillslopes and valley bottoms will develop than the simple linear form required for a V-shaped cross profile.

Typical explanations for the development of U-shaped valleys as a result of glacial erosion rely on the argument either that glacial 'valleys' are actually glacial channels, and that steep side walls and a relatively flat bottom is a characteristic form for fluid flow in channels, or that the cross-sectional pattern of erosion under a glacier includes a wide central maximum leading to steep side walls and a low gradient profile section in the channel centre. Numerical modelling linking ice dynamics, subglacial erosion patterns and cross-profile form development has demonstrated that U-shaped cross sections can result solely from glacial erosion in homogeneous bedrock. However, when spatial variations in rock resistance to erosion are introduced to the model, a wide variety of cross-section shapes can develop, including V-shaped forms. In addition, many glaciated valleys used to illustrate U-shaped valleys or included in morphometric analyses of valley form include substantial

depositional components; the U-shaped form arises from the combination of a low gradient valley floor (fluvioglacial deposition), and a concave talus slope (postglacial and ice marginal slope processes) below steep bedrock walls (glacially modified).

Although the distinction between idealized U-shaped and V-shaped valleys for areas dominated by glacial and fluvial erosion is useful, there is in fact a wide variety of valley cross-profile forms. Other and more complex cross profiles result from temporal and spatial variations in processes across the profile, including both erosion and deposition, and from patterns of surface material resistance to erosion.

Further reading

Augustinus, P.C. (1995) Glacial valley cross-profile development: the influence of *in situ* rock stress and rock mass strength, with examples from Southern Alps, New Zealand, *Geomorphology* 11, 87–97.

Carson, M.A. and Kirkby, M.J. (1972) *Hillslope Form and Process*, Cambridge: Cambridge University Press.

Harbor, J. (1995) Development of glacial-valley cross sections under conditions of spatially variable resistance to erosion, *Geomorphology* 14, 99–107.

Hirano, M. and Aniya, M. (1988) A rational explanation of cross-profile morphology for glacial valleys and of glacial valley development, *Earth Surface Processes and Landforms* 13, 707–716.

SEE ALSO: hillslope, form; hillslope, process; valley

JON HARBOR

CRUSTAL DEFORMATION

Motions of the lithosphere disrupt and modify rocks and the topographic surface. As a manifestation of PLATE TECTONICS, these deformations maintain continental forms that protrude above sea level. Crustal deformations, such as fault offsets and folds, produce diverse constructional landforms dependent on local material properties and surface processes. ACTIVE MARGINS are shaped by competition between deformation and erosion. Deformation occurs at the timescale of plate motions (centimetres per year) but can be slower along individual structures. Recent technologies have revolutionized crustal deformation studies, such as space-based geodesy (e.g. GPS) and seismology that constrain short-term behaviour, dating techniques (e.g. COSMOGENIC DATING) that constrain chronologies of offset geologic markers, and DIGITAL ELEVATION MODELS that permit topographic assessment of large areas.

Crustal deformation leads to OROGENESIS and basin formation over the long term, producing wholesale surface uplift, DENUDATION, and SUBSIDENCE (see TECTONIC GEOMORPHOLOGY). Fluvial systems respond to perturbations in BASE LEVEL where the crust has risen or fallen (Burbank and Anderson 2001). Long-profiles (see LONG PROFILE, RIVER) of stream channels adjust to uplift via KNICKPOINT migration and incision, often leaving behind suites of terraces. Drainage networks may also be modified, as streams can be deflected by zones of uplift or forced to migrate by tilting (see ASYMMETRIC VALLEYS). Sediment loading and gradient changes further influence fluvial form, such as the occurrence of meandering versus BRAIDED RIVER channels. Adjustment to base level in turn affects hillslope processes, leading to increased RELIEF, hillslope length and sediment production. Glacial and coastal erosion similarly respond to uplift and subsidence. Displaced geomorphic features, such as river terraces (see TERRACE, RIVER), shorelines, ALLUVIAL FANS, strata, MORAINES and PLANATION SURFACES, serve as markers that are valuable constraints on relative uplift.

Crustal deformation is most commonly associated with faults (see FAULT AND FAULT SCARP). Dislocations occur along lengths of faults during rupture events, producing earthquakes as a side effect. Ruptures that break the surface are

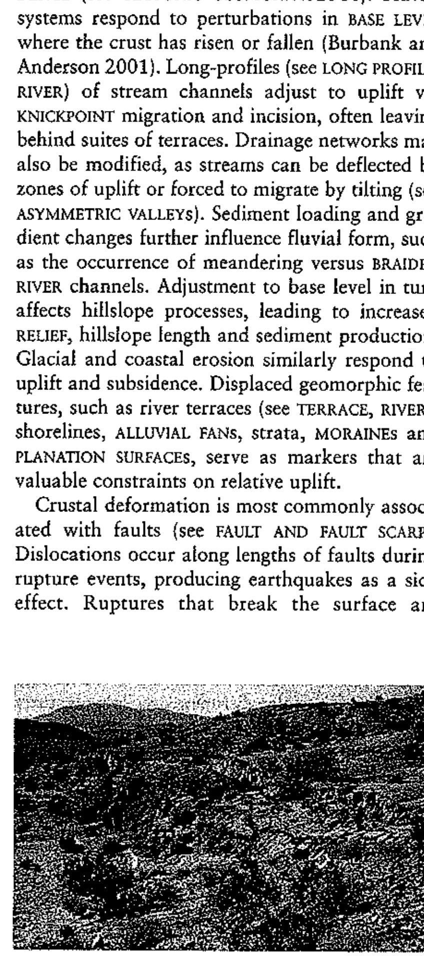

Plate 30 Crustal deformation in alluvium produced locally along the Emerson fault during the 28 June 1992 Landers earthquake in California ($M = 7.3$). The scarp faces to the south-west and is approximately 1 m high. Its height is locally accentuated by lateral offset of the hilly topography. Dextral offset of ~5 m is evident in the displaced stream course. This photograph was taken several days after the earthquake by Kerry Sieh (California Institute of Technology, USA)

typically tens of kilometres long and involve metres of slip. They are quantified in terms of seismic moment: $M_0 = \mu AD$, where μ is rigidity, A is rupture area, and D is average displacement. Earthquake size is thus partly dependent on rupture length, which is controlled by fault zone geometry and segmentation (Plate 30). Coseismic displacement also scales with rupture length, such as the tendency for slip to be $\sim 10^{-4} - 10^{-5}$ of the length of strike-slip fault ruptures. This scaling is related to the elastic strain the crust adjacent to faults sustains during interseismic periods. The release of accumulated strain provides for moderately regular rupture recurrence. Short-lived faulting events are thus the building blocks by which plate motion translates into long-term deformation (Yeats *et al.* 1997). Over the long term, fault displacements tend to scale as several per cent of the total fault length.

Each of the three main types of plate boundaries consists of faults characterized by certain landforms. Strike-slip faults produced by simple shear involve mainly horizontal displacement and create a minimal degree of topographic disruption. Linear troughs are common along such faults, where weakened fault rocks (see CATACLASIS) are easily eroded by deflected stream courses (e.g. the San Andreas fault). Landforms produced by transpression and transtension at restraining and releasing fault bends include pressure ridges, pull-apart basins (see PULL-APART AND PIGGY-BACK BASIN), and variably faced scarps (scissoring). Strike-slip faults also disrupt geomorphic features horizontally, creating shutter ridges (topographic steps) and deflected or BEHEADED VALLEYs and streams (Sieh and Jahns 1984).

Dip-slip faults involve primarily vertical motion. Normal faults are produced by horizontal extension, where maximum compressive stress is oriented vertically. Resulting fault planes typically dip steeply (~60°). Normal faults juxtapose tilted basement blocks and alluvial valleys in the characteristic basin and range terrain. Vertical separation tends to be asymmetric, with valley subsidence exceeding uplift of basement blocks. Edges of uplifted blocks may preserve FLAT IRONS (triangular facets) related to the fault surface. Mountain fronts typically consist of linear segments interrupted by complex transfer zones, such as the Wasatch front (Machette *et al.* 1992). Parallel normal faults produce down-dropped rift valleys (grabens) (see RIFT VALLEY AND RIFTING) and upthrown blocks (HORSTs).

Reverse or thrust faults are produced by horizontal compression, where the least principal stress is oriented vertically. Thrust fault planes dip shallowly (~30°) and produce irregular mountain fronts that involve wide belts of deformation (Philip and Meghraoui 1983). The degree to which such piedmonts are dominated by erosion, deposition and deformation is represented by numerous geomorphic characteristics, including sinuosity, fan entrenchment and valley geometry. Thrust belts typically involve overlapping arcuate fault segments in parallel series that are connected by secondary structure. These may also involve folding, as typical of foreland fold and thrust belts such as along the Nepal Himalaya (Schelling and Arita 1991). Megathrusts of subduction zones create unique cycles of elastic uplift and subsidence in both hanging wall and footwall, leading to rhythmic perturbation of coastal geomorphology.

Deformation along faults during rupture events can be complex. Fault traces tend to be irregular, such as the characteristic en echelon, anastomosing arrangement of faults within wide (~50 m) ruptures of strike-slip faults (Yeats *et al.* 1997). These shear zones can involve pervasive shearing, although slip tends to concentrate along principal displacement zones. A variety of microgeomorphic features are produced during surface ruptures (see SEISMOTECTONIC GEOMORPHOLOGY). Fault scarps record the vertical separation along faults and portray characteristics linked with fault orientation. Scarp degradation through time occurs predictably by incision and diffusive hillslope creep, such that scarp form is related to scarp age (Avouac and Peltzer 1993). These distinctive landforms record deformation history that can be unravelled using palaeoseismology.

Tectonic strain is also accommodated by FOLDing of rock and sediment, particularly in deep basins. Folding of near surface involves permanent brittle deformation in the form of penetrative intergranular shear or flexural slip between strata. Folds are often associated with blind thrust faults and evolve as faults propagate towards the surface. Fold geometry is closely linked with fault bend and tip geometry. Ongoing deposition around folds can result in piggy-back basins and growth strata that itself becomes folded. Erosion and deposition can also mask the topographic expression of folding in unconsolidated sediment. Processes of diagenetic and pedogenic lithification are thus important for fold preservation. Because strata vary in composition and resistance to erosion, ancient folds can be exhumed by erosion, such as palaeo-folds of the Appalachian Valley and Ridge.

References

Avouac, J.-P. and Peltzer, G. (1993) Active tectonics in southern Xinjiang, China: analysis of terrace riser and normal fault scarp degradation along the Hotan-Qira fault system, *Journal of Geophysical Research* 98, 21,773–21,807.

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

Machette, M.N., Personius, S.F. and Nelson, A.R. (1992) The Wasatch Fault Zone, U.S.A., *Annales Tectonicae* 6, 5–39.

Philip, H. and Meghraoui, M. (1983) Structural analysis and interpretation of the surface deformations of the El Asnam earthquake of October 10, 1980, *Tectonics* 2, 17–49.

Schelling, D. and Arita, K. (1991) Thrust tectonics, crustal shortening, and the structure of the far-eastern Nepal Himalaya, *Tectonics* 10, 851–862.

Sieh, K.E. and Jahns, R.H. (1984) Holocene activity of the San Andreas fault at Wallace Creek, California, *Geological Society of America Bulletin* 95, 883–896.

Yeats, R.S., Sieh, K. and Allen, C.R. (1997) *The Geology of Earthquakes*, Oxford: Oxford University Press.

JAMES A. SPOTILA

CRUSTING OF SOIL

Crusts are thin layers, different in character from the soil beneath, that develop at the interface between the soil and the atmosphere.

One class of crust is often termed inorganic or rain-beat crusts. Large amounts of energy and high transient forces are imparted to the surface of the soil by the impact of raindrops. These break down soil aggregates, compress the surface and dislodge particles. This physical disruption, which may be especially effective where vegetation cover is limited, is aided by certain chemical processes, which include dispersion, which cause further breakdown of soil aggregates. The resulting dense surface layer forms a surface seal, and when this seal dries, a crust forms. Such a crust can have profound effects on runoff and on erosion by wind and water (Poesen and Nearing 1993).

Recently it has been recognized that organic (also called microphytic, microbiotic, cryptogamic or biological) crusts in and on the surfaces of soils play important hydrological and geomorphological roles (Eldridge and Rosentreter 1999). Organic compounds, including plant waxes, can produce hydrophobic (water repellent) substances, as can a range of fungi and soil micro-organisms. Although water repellent soils occur in more humid environments, there are many examples of them that have been reported from semi-arid areas (Doerr *et al.* 2000). These hydrophobic surfaces tend to be zones of reduced soil infiltration capacity and thus of increased overland flow. Following from this is the likelihood that enhanced soil erosion also occurs. Removal of the crusts has been shown to have a dramatic effect on infiltration rates (Eldridge *et al.* 2000).

Likewise biological soil crusts have an influence on aeolian processes. A cover of cyanobacteria, green algae, lichens and mosses is important in stabilizing soils in drylands and thus protects them from wind erosion. They play a role in dune stabilization (Kidron *et al.* 2000). Unlike vascular plants, the cover of organic crusts is not reduced in drought years and they are present the whole year round. However, they are very susceptible to anthropogenic disturbance (Belnap and Gillette 1997). Filamentous cyanobacteria mats are especially effective against wind attack (McKenna-Neuman *et al.* 1996). The filaments and extracellular secretions of cyanobacteria also form water stable aggregates that help soils to resist water erosion and raindrop impact effects (Issa *et al.* 2001). It also needs to be appreciated that not all organic crusts are hydrophobic and that by eliminating the effect of raindrop impact, they prevent the rapid development of a sealed rain-beat crust conducive to runoff generation (Kidron and Yair 1997).

References

Belnap, J. and Gillette, D.A. (1997) Disturbance of biological soil crusts: impacts on potential wind erodibility of sandy desert soils in southeastern Utah, *Land Degradation and Development* 8, 355–362.

Doerr, S.H., Shakesby, R.A. and Walsh, R.P.D. (2000) Soil water repellancy: its causes, characteristics and hydro-geomorphological significance, *Earth-Science Reviews* 51, 33–65.

Eldridge, D.J. and Rosentreter, R. (1999) Morphological groups: a framework for monitoring microphytic crusts in arid landscapes, *Journal of Arid Environments* 41, 11–25.

Eldridge, D.J., Zaady, G. and Shachack, M. (2000) Infiltration through three contrasting biological soil crusts in patterned landscapes in the Negev, Israel, *Catena* 40, 323–336.

Issa, O.M., Le Bissonnais, Y., Défrage, C. and Trichet, J. (2001) Role of cyanobacterial cover on structural stability of sandy soils in the Sahelian part of Western Niger, *Geoderma* 101, 15–30.
Kidron, G.J. and Yair, A. (1997) Rainfall-runoff relationship over encrusted dune surfaces, Nizzama, western Negev, Israel, *Earth Surface Processes and Landforms* 22, 1,169–1,184.
Kidron, G.J., Barzilay, E. and Sachs, E. (2000) Microclimate control upon sand microbiotic crusts, western Negev Desert, Israel, *Geomorphology* 36, 1–18.
McKenna-Neuman, C., Maxwell, C.D. and Bouton, J.W. (1996) Wind transport of sand surface crusted with photoautotrophic micro organisms, *Catena* 27, 229–247.
Poesen, J.W.A. and Nearing, M.A. (eds) (1993) Soil surface sealing and crusting, *Catena Supplement* 24.

A.S. GOUDIE

CRYOPLANATION

Cryoplanation (Bryan 1946) is a morphogenetic term introduced to explain and describe low-angled slope surfaces occurring on higher valley-side and summit positions (cryoplanation terraces, or benches), or in valley-side foot positions (cryopediments) in periglacial regions. Cryoplanation and altiplanation are synonymous, and both are forms of equiplanation. Cryoplanation terrace has subsumed several other periglacial terrace terms, including: goletz, altiplanation, NIVATION and equiplanation.

Alleged cryoplanation terraces have flats or treads of 1° to 12° with a sharp inflection in slope at the upslope limit (sometimes called the knickpoint) where risers are often 25° to 35°. Terrace width is often only a few metres, but claims in excess of one kilometre exist (Demek 1969). Sets of cryoplanation terraces produce a staircase effect on a hillside, and their convergence on a summit from two or more sides may produce a summit flat. Both terrace size and frequency appear to increase with time since deglaciation, but terraces may also occur in unglaciated regions. Terrace relationship to permafrost and rock structure is extremely uncertain, but adjustment to rock type is reported. Transport of debris across entire sets of cryoplanation terraces seems essential in some circumstances. This appears to be problematic as lower terraces would have to export all debris from upslope terraces unless it was shed laterally which seems unlikely.

Cryopediments are subject to the same uncertainties associated with tropical pedimentation. A cryopediment is viewed as expanding by headward incision by freeze–thaw weathering, or nivation more broadly. The flat is viewed as a bedrock surface veneered by debris experiencing common periglacial mass wasting processes, e.g. SOLIFLUCTION.

As a process cryoplanation has no unique elements but appears to be synonymous with nivation (Thorn and Hall, 2003) which itself merits more precise articulation. While emphasis has been placed on nivation during the early stage of cryoplanation specifically (Demek 1969), no other specific process (while implied) has ever been invoked for the mature stages. If large perennial snowpatches are protective rather than erosive, as well may be the case, largeness and/or increasingly cold climate may not favour headward expansion. The presence of patterned ground on the tread or transport surface is often, but not always, invoked as an indicator of inactivity.

While the landforms designated cryoplanation terraces or cryopediments are clearly found in periglacial environments, the general absence of sound process research (but see Hall 1997) renders their origins unknown. This problem is exacerbated by the apparent present inactivity of many such features. Nelson (1989) has suggested that cryoplanation terraces may be a periglacial analogue of cirque glaciation reflecting a precipitation/temperature regime unable to sustain full glaciation. Hall (1998) and Thorn and Hall (2003) have suggested that the distinction between cryoplanation and nivation forms and processes needs careful re-examination as it is presently far from clear.

References

Bryan, K. (1946) Cryopedology – the study of frozen ground and intensive frost-action with suggestions on nomenclature, *American Journal of Science* 244, 622–642.
Demek, J. (1969) Cryoplanation terraces, their geographical distribution, genesis and development, *Československé Akademie Věd Rozpravy, Řada Matematických a Přírodních Věd*, 79(4).
Hall, K. (1997) Observations on 'cryoplanation' benches in Antarctica, *Antarctic Science* 9, 181–187.
——(1998) Nivation or cryoplanation: different terms, same features? *Polar Geography* 22, 1–16.
Nelson, F.E. (1989) Cryoplanation terraces: periglacial cirque analogs, *Geografiska Annaler* 71A, 31–41.
Thorn, C.E. and Hall, K. (2003) Nivation and cryoplanation: the case for scrutiny and Integration, *Progress in Physical Geography* 26, 553–560.

Further reading

Priesnitz, K. (1988) Cryoplanation, in M.J. Clark (ed.) *Advances in Periglacial Geomorphology*, 49–67, Chichester: Wiley.

SEE ALSO: nivation

COLIN E. THORN

CRYOSTATIC PRESSURE

The elevated water potential in saturated, coarse-grained sediments caused by freezing in a closed system. Where the volumetric expansion of water by 9 per cent on becoming ice cannot be accommodated in freezing sediment, pore water is expelled into proximal unfrozen ground, raising the water pressure. Cryostatic pressure is responsible for the uplift of closed-system pingos, beneath which pressures of up to 0.4 MPa have been measured (Mackay 1977). In fine-grained soils, cryostatic pressure may develop at the beginning of laboratory freezing tests, but it has not been measured under field conditions.

Reference

Mackay, J.R. (1977) Pulsating pingos, Tuktoyaktuk Peninsula, N.W.T., *Canadian Journal of Earth Sciences* 14, 209–222.

C.R. BURN

CRYPTOKARST

Cryptokarst is a form of karstification limited to the EPIKARSTIC zone. It is always developed under a cover of superficial formations resulting from deposition (loess, etc.) or weathering (alterite). The quality and the thickness of the superficial formation have a direct influence on cryptokarst activity (Nicod 1994).

The main source of acid in ground water is the surface, with the percolation of humic acid (biological activity) and the gaseous exchanges between atmosphere and rainwater. Thus the epikarstic zone is submitted to intense dissolution (Klimchouk 1995) due to the proximity of the surface. The formation of the cryptokarst is enabled by the layer of superficial formation that distributes the water in a diffuse manner, avoiding the concentration of water with high dissolution potential. The chemical equilibration between water and terranes (Stumm and Morgan 1981) implies that the dissolution capacity will decrease proportionally to the residence time of water in the epikarst zone. To stay in the cryptokarst phase, the karstification has to be aborted before reaching the underlying rocks. It means that there will not be any transmission of aggressive water below the epikarst zone, i.e. no water at all either because there are no fast paths (diaclases) or water is non-aggressive because it has already reached equilibrium with the rocks. Concentration of clayey particles that originate from the weathering of the superficial formations can also lead to the clogging of the bedrock interface. In some circumstances, the superficial formations may be drawn down with the vertical progression of the cryptokarstic front and this may induce surface depressions like dolines. On the other hand, the layer of superficial material protects the cryptokarst from surface mechanical erosion.

The geological and topographical conditions for cryptokarst are found in the chalky Cretaceous formations of the Paris Basin (Rodet 1992), England and Denmark. The chalky basement is slightly tectonized (with a resultant low density of fracturation) and the relief is composed of plateaus separated by DRY VALLEYS. The carbonate components of the chalk are easily dissolved but the argillaceous part remains in place. The argillaceous particles are hardly removed by the horizontal water movement. This causes a reduction of the permeability and of the capacity of the basement to be eroded (Lacroix *et al.* 2002).

References

Klimchouk, A.B. (1995) Karst morphogenesis in the epikarstic zone, *Cave and Karst Science* 21, 45–50.
Lacroix, M., Rodet, J., Wang, H.Q., Laignel, B. and Dupont, J.P. (2002) Microgranulometric approach to a chalk karst, western Paris Basin, France, *Geomorphology* 44, 1–17.
Nicod, J. (1994) Plateaux karstiques sous couverture en France, *Annales de Géographie* 576, 170–194.
Rodet, J. (1992) *La craie et ses karsts*, Caen: Ed. CNEK-Groupe Seine.
Stumm, W. and Morgan, J.J. (1981) *Aquatic Chemistry*, New York: Wiley.

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

MICHEL LACROIX

CRYPTOVOLCANO

A roughly circular area of greatly disturbed rocks and sediments that is morphologically suggestive of being the result of volcanic activity but does not contain any true volcanic materials. Very often the origin of the features has been a matter of controversy. For example, the Pretoria Salt Pan crater in South Africa and the great Vredefort Dome have in the past sometimes been interpreted as volcanic features, but there is now an accumulation of evidence that they both result from meteorite impact (Reinold *et al.* 1992; Reinold and Coney 2001). Conversely, Upheaval Dome in the Canyonlands National Park, Utah, USA, has variously been attributed to meteoric impact, fluid escape, cryptovolcanic explosion and salt doming, with the last explanation now being favoured (Jackson *et al.* 1998). Some features, including a group of eight structures running in a 200-km straight line across the USA, may be the result of comet or asteroid impact (Rampino and Volk 1996) (see ASTROBLEMES, CRATERS). The cryptovolcanic features discussed above show a great range in size. The Pretoria Salt Pan crater is 1.13 km in diameter, whereas the Vredefort structure is 250–300 km across. The aligned structures in the USA are *c*.3–17 km in diameter.

The problem of establishing the origin of closed depressions and circular structures is made evident when one considers the range of hypotheses that have been put forward to explain the Carolina Bays in the eastern USA (Ross 1987):

1 Spring basins
2 Sand bar dams or drowned valleys
3 Depressions dammed by giant sand ripples
4 Craters of meteor swarm
5 Submarine scour by eddies, currents or undertow
6 Segmentation of lagoons and formation of crescentic keys; original hollows at the foot of marine terraces and between dunes
7 Lakes in sand elongated in direction of maximum wind velocity
8 Solution depression, with wind-drift sand forming the rims
9 Solution depressions, with magnetic highs near bays due to redeposition of iron compounds leached from the basins
10 Basins scoured out by confined gyroscopic eddies
11 Solution basins of artesian springs with lee dunes
12 Fish nests made by giant schools of fish waving their fins in unison over submarine artesian springs
13 Aeolian blowouts
14 Bays are sinks over limestone solution areas streamlined by ground water
15 Oriented lakes of stabilized grassland interridge swales of former beach plains and longitudinal dunefields with some formed from basins in Pleistocene lagoons
16 Black hole striking in Canada (Houston Bay) throwing ice onto coastal plain
17 Cometary fragments exploding above surface, their shock waves creating depressions
18 Drought with subsequent fire in peat bogs followed by aeolian activity.

References

Jackson, M.P.A, Schultz-Ela, D.D., Hudec, M.R., Watson, I.A. and Porter, M.L. (1998) Structure and evolution of Upheaval Dome: a pinched-off salt diapir, *Geological Society of America Bulletin* 110, 1,547–1,573.

Rampino, M.R. and Volt, T. (1996) Multiple impact event in the Paleozoic: collision with a string of comets or asteroids? *Geophysical Research Letters* 23, 49–52.

Reinold, W.V. and Coney, L. (2001) *The Vredefort Impact Structure and Directly Related Subjects: An Updated Bibliography*, Economic Geology Research Institute, University of the Witwatersrand, Johannesburg, Information Circular No. 353.

Reinold, W.V., Koeberl, C., Partridge, T.C. and Kerr, S.J. (1992) Pretoria Saltpan crater: impact origin confirmed, *Geology* 20, 1,079–1,082.

Ross, T.E. (1987) A comprehensive bibliography of the Carolina Bays Literature, *Journal of the Elisha Mitchell Scientific Society* 103, 28–42.

A.S. GOUDIE

CUESTA

An asymmetric ridge built of dipping sedimentary rocks of alternating resistance against weathering and erosion, elongated along the strike of strata, is called a cuesta. The steep front slope is opposite to the dip, whereas the gently sloping backslope is more or less parallel to the dip. The top part of the cuesta face and the backslope are built of more resistant strata; less resistant ones are exposed in the lower part of the front scarp.

Because of contrasting slope and lithology, each side of a cuesta is shaped by different sets of

processes. Rapid mass movement and gully erosion predominate on the steeper slope, and fluvial incision and slow mass movement operate on the backslope. Hence in the long term a cuesta both retreats and is worn down. There is a number of theories how cuesta ridges form but most emphasize differential fluvial erosion within a monocline, which leaves outcrops of more resistant strata as divides and initial cuestas, which then begin to retreat. Bevelled ridge tops indicate that cuesta have developed from a former plain through river incision.

Cuesta is an example of a structure-controlled and climate-independent landform. Classic cuesta landscapes include the Colorado Plateau in North America, the Paris Basin in France and Southern German Uplands.

Further reading

Ahnert, F. (1996) *Einführung in die Geomorphologie*, Stuttgart: Ulmer.

Schmidt, K.-H. (1994) The groundplan of cuesta scarps in dry regions as controlled by lithology and structure, in D.A. Robinson and R.B.G. Williams (eds) *Rock Weathering and Landform Evolution*, 355–368, Chichester: Wiley.

SEE ALSO: caprock; escarpment; mesa; sandstone geomorphology; structural landform

PIOTR MIGOŃ

CURRENT

The hydrodynamics responsible for sediment entrainment and transport, erosion and accretion, and morphological change within the coastal zone, consist of *oscillatory motions* associated with WAVES of various frequencies and forms and *quasi-steady, unidirectional currents.* The currents are forced by: (1) a secondary effect of the waves themselves, i.e. *wave drift* or *wave streaming*; (2) *tides*; (3) *wind* stress; (4) *pressure* and *density* gradients; and (5) a variety of motions resulting from the dissipation of wave energy at, and landward of, depth-controlled breaking (*surf zone*). Here the kinetic energy of the waves is transformed into: (a) increased macro and micro turbulence; (b) *drift currents* associated with secondary progressive or standing waves, generally of lower frequency than the incident waves (e.g. edge waves, leaky waves, etc.); (c) *longshore currents*; (d) *rip currents*; (e) *undertow*; (f) *swash*.

Wave streaming (wave drift)

Stokes in 1847 first recognized that WAVE orbital motions were not closed in the case of small amplitude waves in a perfect non-viscous fluid, even in deep water. The fluid particles have a second-order, wave-averaged, mean *Lagrangian* velocity and thus there is a finite *mass flux* of water. Since horizontal velocities increase slightly with distance above the bed, so that particle motion under the crest is slightly larger than under the trough, conservation of mass causes a stratification of flow (Figure 29a). In shallow water, with greater bed friction, the wave orbits become elliptical and drift velocities increase ($\sim 0.1\,m\,s^{-1}$). A *Eulerian* measure of mass flux can also be obtained by integrating the horizontal velocities beneath the crest and trough over space and time; the same mass flux is obtained although the vertical distribution is different. For real viscous fluids, and waves in finite depth, Longuet-Higgins (1953) showed that there is a time-averaged, net downward transfer of momentum into the boundary layer at the bed, producing a *Eulerian* streaming in addition to the Stokes drift. Again by conserving mass, a stratified profile of the mean current is obtained (Figure 29b); flow is in the direction of propagation at the bed and a reversal occurs at mid-depth; Klopman (1994) has confirmed this pattern through laboratory experiments. In strongly asymmetric flows over steep slopes, shear stresses within the boundary layer may cause a reversal (upwave) mean current at the bed.

Surf zone currents

Currents in the surf zone interact with the instantaneous wave orbital motions (over the full range of short and long period waves) producing a rather complex time-dependent three-dimensional pattern (Svendsen and Lorenz 1989; Figure 30). This is usually disaggregated into a number of distinct components:

Longshore currents are generated when waves break at an oblique angle to the shoreline, and the alongshore component of the onshore directed

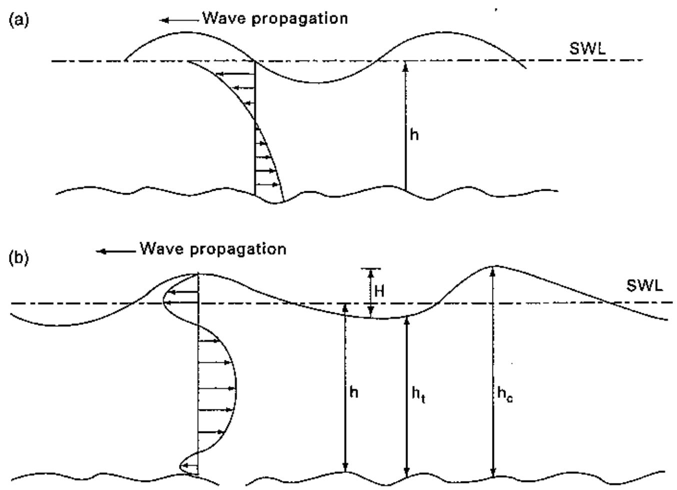

Figure 29 Wave-induced quasi-steady currents: (a) Stokes drift; (b) Longuet-Higgins mass transport

radiation stress (see WAVE) forces a shore-parallel or *longshore current*. Pressure gradients due to water level *set-up* differentials along shore, as well as the shore-parallel component of the onshore *wind stress*, can enhance (or reduce) this forcing. Laboratory and field measurements indicate that the longshore current increases landward from the breakpoint, reaches a maximum around the mid-surf zone and decreases to near zero at the shoreline. Since the radiation stress gradient is a maximum at the breakpoint in the ideal theoretical solution (Longuet-Higgins 1970a,b), a 'smoothing' of the momentum flux across the surf zone, called *lateral mixing*, causes the maximum current to be displaced landward. This mixing also causes *longshore currents* to flow outside the zone of breaking, even though the radiation stress gradients approximate zero. Komar (1998) gives a detailed review of the origins and the spatial and temporal patterns of longshore currents.

Undertow or *near-bed return flow* is a pressure gradient driven, time-averaged, mean current directed seawards near to the bed, everywhere along the shoreline. It is caused by cross-shore differences in the mean water elevation due to wave *set-up* at the shoreline and *set-down* under the breaker zone. *Set-up* and *set-down* result from differences in the local *onshore flux of momentum* by waves (*radiation stress*), which is largest at the breakpoint (where waves are largest) and smallest at the final point of wave dissipation at the shoreline. This gradient forces a displacement of the water from beneath the largest waves towards the shoreline and will be complemented by the onshore *mass flux* of water by the waves, as well as any water moved by *wind stress* acting towards the shoreline. Where nearshore sand bars are present, multiple set-ups and set-downs and associated undertows may be formed by the multiple breaker lines (Greenwood and Osborne 1990). Typically velocities are small, but recordings have been made of undertows up to $0.80\,m\,s^{-1}$.

Rip currents are discrete, narrow, high velocity *jets* of offshore-directed flow across the surf zone, often forming part of a regular horizontal cellular circulation, with associated shore-parallel to oblique *feeder currents*, and an area of flow expansion, the *rip head*, seaward of wave breaking (Figure 31). *Rips* are often associated with cross-shore oriented depressions (*rip channels*) or breaks in a quasi-shore parallel bar, but are found also on uniformly sloping beaches. Rip currents are not generally *steady state* phenomena, but vary both spatially and temporally.

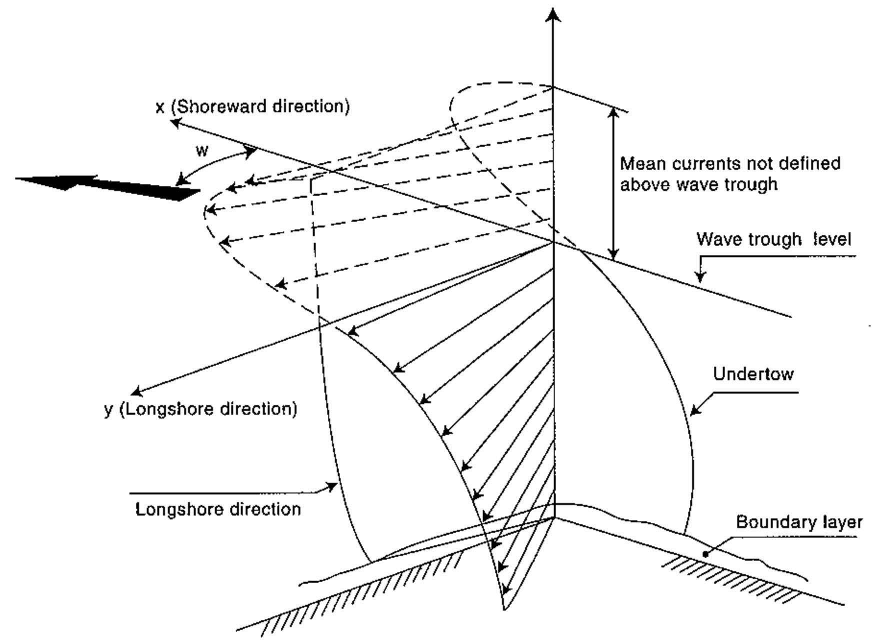

Figure 30 Time-averaged mean velocity vectors in the surf zone (modified after Svendsen and Lorenz 1989)

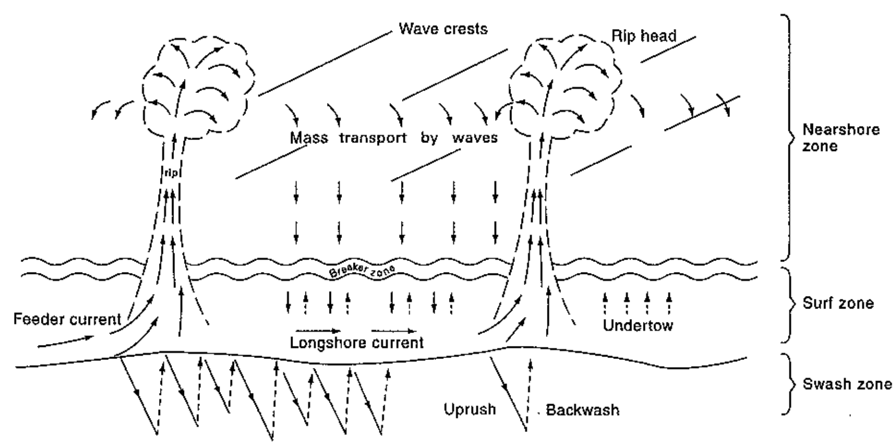

Figure 31 Horizontal cellular circulations in the surf zone

They may be spatially periodic alongshore, with spacing ranging from a few metres (*micro-rips* associated with the embayments of *beach cusps*), to order of 10^2 m (storm cusps), to *mega rips* (Short 1985), which are often single, large-scale, topographically controlled *jets* induced by cellular circulation within the confines of headlands/structures. *Mega rips* may flow offshore for more than one kilometre and can reach speeds $> 3\,m\,s^{-1}$. Theoretically, rip currents result from a periodic modulation of wave heights and thus *wave set-up* alongshore. This modulation has been related to: (1) rhythmic variations in topography (Sonu 1972; Komar 1971); (2) edge waves (Bowen 1969; Bowen and Inman 1969); (3) interference between two incident wave fields (Dalrymple 1975). Rip currents typically pulse at infragravity wave frequencies (< 0.04 Hz), most likely as a result of the increased *radiation stresses* at the *wave group* frequency. Aagaard *et al.* (1997) measured very long period fluctuations of 5–10 minutes. Current speeds may increase with falling tidal levels, especially where topography increasingly confines the current. Rip currents are capable of transporting large volumes of both coarse bedload (especially in migrating megaripples, e.g. Gruszczynski *et al.* 1993) and suspended load from within the surf zone to seaward. Mega rips may well transport nearshore sediments onto the continental shelf below the level of average *wave base* (depth limit of surface waves).

Swash is the *uprush* and *backwash* currents on the beach face and reflects the ultimate dissipation of incident wave energy. Both currents are turbulent, with the former assuming a flow direction coincident with the angle of approach of the breaking wave; the backwash results from gravity acting on the water on the beach and thus flows down the maximum beach slope. This often results in a 'zig-zag' motion of water and sediment. Swash currents depend on the nature of the incoming waves, the beach face slope and the state of the beach water table. If the beach is not saturated there will be a tendency for infiltration of the uprush into the permeable beach face, and thus a reduction in the amount of water in the backwash. Because water can drain from the beach face water table, the backwash tends to last longer and is usually thinner and flows may be supercritical, with hydraulic jumps common. Van Rijn (1998) and Butt and Russell (2000) provide reviews of *swash* hydrodynamics and sediment transport.

Tidal currents or tidal streams

Currents that result from the tidal wave forced by the gravitational tractive stresses generated by the moon and sun are called *tidal currents* or *tidal streams* and reverse their direction either semi-diurnally or diurnally. They may reach speeds up to $6\,m\,s^{-1}$ in coastal waters if they are constrained topographically, but are generally much smaller ($\sim 0.05\,m\,s^{-1}$) and are dominated on open coasts by gravity WAVE oscillations. Tidal currents vary in magnitude in response to the local tidal range and will have a variable phase relationship with tidal elevation depending upon whether the tidal wave is *standing* (maximum flows at mid-tide) or *progressive* (maximum flows at high and low tide). The current direction will be constrained by Coriolis and thus currents generated by the rising (flood) tide will follow a different path than those of the falling (ebb) tide, giving an ellipsoidal pattern of currents. In estuaries, for example, distinct *flood* and *ebb channels* may exist. This creates *residual currents* that may be significant in terms of sediment transport and deposition. In all cases tidal currents vary with depth as a result of bottom friction and a logarithmic velocity profile develops. Tidal currents are most significant in estuaries, inlets and straits and in some sections of continental shelf. Davis and Hayes (1984) discuss the relative role of tides and waves in the development of coastal morphologies.

Other currents in the coastal ocean

Wind-induced currents are formed when wind shear on the surface is transferred into the water column. At the shoreline this will induce a mass transport in the direction of the wind and can be resolved into a shore-normal and shore-parallel component. The former results in an elevation of the mean sea level (*wind set-up*) at the shoreline, which is an addition to the *wave set-up* which causes *undertows*; the latter will enhance the radiation stress driven *longshore current*. However, such currents are often short-lived, as local winds are subject to frequent change in speed and direction. Gradient in *wind set-up* can also generate currents at the scale of the complete coastal ocean boundary layer during large storm events (e.g. hurricanes, intense mid-latitude cyclones). This results in large offshore directed pressure-gradient flows, whose speed and direction are constrained by frictional forces and the Coriolis effect (Swift 1976). In deep water and where wind systems are of long duration (e.g. the Trade Winds, Equatorial Winds, other zonal winds, etc.) they cause large coastal circulation systems (e.g. upwelling and downwelling systems, the Equatorial currents, etc.). The rotational force of the moving Earth (Coriolis) also influences such large-scale flows. Currents tend to move at 45 degrees to the wind at the surface and rotate clockwise (or anticlockwise depending on the hemisphere) with depth, to flow in the opposite direction to the surface wind; this is the *Ekman Spiral*.

Density-induced currents result from density differences due to differences in temperature, salinity or sediment mass concentration. Such gradients force both horizontal and vertical currents, which are also affected by the rotational effect of the Earth. The component of flow driven by the slope of an internal density surface is called the *baroclinic* component; the component driven by the slope of the sea surface is the *barotrophic* component.

Inertial currents are residual currents in large bodies of water, which continue to flow under their own momentum long after the original forcing has ceased.

References

Aagaard, T., Greenwood, B. and Nielsen, J. (1997) Mean currents and sediment transport in a rip channel, *Marine Geology* 140, 25–25.
Bowen, A.J. (1969) The generation of longshore currents on a plane beach, *Journal of Marine Research* 27, 206–214.
Bowen, A.J. and Inman, D.L. (1969) Rip currents, II Laboratory and field observations, *Journal of Geophysical Research* 74, 5,479–5,490.
Butt, T. and Russell, P. (2000) Hydrodynamics and cross-shore sediment transport in the swash zone of natural beaches: a review, *Journal of Coastal Research* 16, 255–268.
Dalrymple, R.A. (1975) A mechanism for rip current generation on an open coast, *Journal of Geophysical Research* 80, 3,485–3,487.
Davis, R.A. Jr. and Hayes, M.O. (1984) What is a wave-dominated coast, *Marine Geology* 60, 313–329.
Greenwood, B. and Osborne, P.D. (1990) Vertical and horizontal structure in cross-shore flows: an example of undertow and wave set-up on a barred beach, *Coastal Engineering* 14, 543–580.
Gruszczynski, M., Rudowski, S., Semil, J., Slominski, J. and Zrobek, J. (1993) Rip currents as a geological tool, *Sedimentology* 40, 217–236.
Klopman, G. (1994) Vertical structure of flow due to waves and currents, *Report H 840.30*, Delft Hydraulics, Delft, The Netherlands.
Komar, P.D. (1971) Nearshore circulation and the formation of giant cusps, *Geological Society of America Bulletin* 82, 2,643–2,650.
——(1998) *Beach Processes and Sedimentation*, Upper Saddle River, NJ: Prentice Hall.
Longuet-Higgins, M.S. (1953) Mass transport in water waves, *Philosophical Transactions of the Royal Society of London, Series A* 245, 535–581.
——(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.
Short, A.D. (1985) Rip current type, spacing and persistence, Narrabeen Beach, Australia, *Marine Geology* 65, 47–71.
Sonu, C.J. (1972) Field observations of nearshore circulation and meandering currents, *Journal of Geophysical Research* 77, 3,232–3,247.
Svendsen, I.A. and Lorenz, R.S. (1989) Velocities in combined undertow and longshore currents, *Coastal Engineering* 13, 55–79.
Swift, D.J.P. (1976) Coastal sedimentation, in D.J. Stanley and D.J.P. Swift (eds) *Marine Sediment Transport and Environmental Management*, 255–310, New York: Wiley.
Van Rijn, L.C. (1998) *Principles of Coastal Morphology*, Amsterdam: Aqua Publications.

BRIAN GREENWOOD

CUSPATE FORELAND

Cuspate forelands are large-scale, tooth-shaped coastal promontories. Although erosion plays a part in their evolution and their form, they are basically landforms of accretion, composed of sorted beach sand deposited from littoral transport (see LONGSHORE (LITTORAL) DRIFT). They often enclose a lagoon (see LAGOON, COASTAL) or marsh. There are two basic types.

Recurved cuspate forelands

At sites where the coastline changes direction abruptly landward, littoral transport slows and deposition occurs, creating over time, a broad elongated shoal. This feature, termed 'spit platform' by Meistrell (1966), is the foundation on which the emergent cuspate foreland grows. The original elongated form is sometimes referred to as a 'flying spit', 'spit with recurves' or 'fleche', in that its growth is in a direction continuous with that of the updrift coast, 'flying' offshore into deeper water. On the leeward side, sand deposits washed over during storms or transported around the tip are subjected to wave action from the opposite direction. The result is a series of concave-seaward recurves, or secondary spits, extending at an acute angle from the tip to the downstream coast. Because of the effect of this bi-directional wave climate, the foreland may range in form from symmetrically cuspate, when the wave effect is fairly balanced on both sides, to asymmetrical and elongated, if wave effect on one side predominates. Examples of this type are Cape Canaveral in Florida, Pointe de la Coubre near the Gironde estuary in western France, and the Toronto Islands of Lake Ontario, Canada.

Dungeness-type cuspate forelands

This is the term originally given by Gulliver (1895) and Johnson (1919) and elaborated by Zenkovitch (1967), to refer to symmetric, accretionary forelands that grow at high angles to the shore. Coakley (1976) demonstrated that they usually form at the site of a pre-existing morphological feature that is transverse to the coastline, e.g. a recessional moraine, or low bedrock ridge. This disruption of the coastal orientation causes the accumulation of the spit platform. These forelands are dynamic and are influenced by periodic reversals in littoral drift direction due to changes in the wind/wave climate. Thus, the foreland may be nourished, and be eroded, from both sides. This results in the classic pointed cuspate form with a well-developed complex of beach ridges. The evolution of the foreland may be studied through the pattern of the preserved beach ridges. Good examples are Dungeness on the south-eastern English coast and Point Pelee, Lake Erie, Canada.

References

Coakley, J.P. (1976) The formation and evolution of Point Pelee, western Lake Erie, *Canadian Journal of Earth Science* 13(1), 136–144.
Gulliver, F.P. (1895) Cuspate forelands, *Geological Society of America Bulletin* 7, 399–422.
Johnson, D.W. (1919) *Shore Processes and Shoreline Development*, New York: Wiley.
Meistrell, F.J. (1966) The spit-platform concept: laboratory observation of spit development, in M.L. Schwartz (ed.) *Spits and Bars*, 225–284, Stroudsburg, PA: Dowden, Hutchinson and Ross.
Zenkovitch, V.P. (1967) *Processes of Coastal Development*, New York: Interscience Publishers.

SEE ALSO: beach cusp; tombolo

JOHN P. COAKLEY

CUT-AND-FILL

Cut- (or scour) and-fill is the local cyclic erosion and deposition of sediment in a river channel, usually over short time periods (hours to years). It occurs as part of the process of sediment transport and development of channel morphology and consequently is associated with spatial and/or temporal changes in flow conditions, such as the passage of a flood wave along a channel. Cut-and-fill is distinct from progressive changes in channel elevation over longer time spans and greater distances that are usually referred to as degradation (erosion) and aggradation (deposition) and may produce substantial accumulation of ALLUVIUM and formation of terraces (see TERRACE, RIVER).

Cut-and-fill occurs in alluvial stream channels whenever bed sediment is moved. It is the result of variation in channel topography related to the normal processes of channel development, sediment transport and the response to, and recovery from, events such as large floods. Consequently cut-and-fill occurs for a variety of reasons including changes in flow hydraulics and sediment transport rate along the stream or during a flood event, development and migration of BEDFORMS, and changes in channel pattern, position or overall morphology. Cut-and-fill related to changes in channel morphology or channel migration is well known from studies of braided streams (see BRAIDED RIVERS) and is associated with the formation and migration of scour pools and bars (see BAR, RIVER) and with channel migration and AVULSION.

A cycle of cut-and-fill may occur at a single cross section during a flood, sometimes associated with rising and falling stages of the hydrograph. For example, in a POOL AND RIFFLE channel, cutting followed by filling may occur in pools while the reverse occurs in riffles because of changes in velocity and bed shear stress as discharge rises and falls. In other cases there are distinct areas of the channel in which only cut or fill occurs during a flood event or over longer time periods. Generally, there is compensating cut-and-fill within a channel reach so that the quantity of erosion at one location is matched by deposition nearby, sediment may be transferred from one to the other and this mass conservation means that there is no overall change in channel elevation (Colby 1964; Ashmore and Church 1998; Eaton and Lapointe 2001).

Observations in large SAND-BED RIVERS have shown cut, and subsequent fill, of the order of two or three metres at particular river cross sections during a single flow event (Colby 1964). In small, GRAVEL-BED RIVERS and streams, measurements indicate that the average depth of cut-and-fill during sediment transport events is of the order of about twice the maximum grain size, but local depths may be much greater than this (Hassan 1990; Haschenburger 1999). The average and maximum depth of cut or fill in a particular stream tends to be greater at higher discharges (greater bed shear stress) as does the area of the channel experiencing cut or fill. Where cut-and-fill is related to the development and migration of bedforms and scour pools the depth of activity is determined by the vertical amplitude of the channel topography.

Common methods for measurement of cut-and-fill are depth sounding, survey of topographic changes over an area of channel, and deployment of scour chains or tracers. Sounding provides very high temporal resolution but may be limited in spatial coverage while repeated surveying provides detailed information on the spatial pattern but may underestimate cut-and-fill amounts and rates if there is both erosion and deposition at a given location between surveys. Scour chains can provide both spatial patterns and also some information about the alternation of cut-and-fill at a point during a flow event. Scour chains are inserted vertically into the stream bed so that the increase in length of chain exposed at the bed after a flow event indicates the depth of cutting, while the depth of fill can be inferred from the depth of burial of the vertical section of chain.

Cut-and-fill is fundamentally and practically significant. Fundamentally, it is the result of the direct connection between channel morphology and sediment transport – spatial and temporal variation in transport rate leads to cut-and-fill and therefore change in channel morphology. Furthermore, the rate of transport of bed sediment during a transport event can be defined as the average depth of cut or fill multiplied by the average velocity of the sediment particles (distance moved divided by the duration of the transport event), which is one method for estimating bed sediment transport rate (Ashmore and Church 1998; Haschenburger 1999). Because cut-and-fill is a significant aspect of stream channel dynamics it is also important in a number of other contexts such as sedimentological interpretation of alluvial deposits (Best and Ashworth 1997), engineering design of river structures, and anticipation of the effects of direct or indirect modification of river channels on channel dynamics and stream habitat.

References

Ashmore, P. and Church, M. (1998) Sediment transport and river morphology: a paradigm for study, in P.C. Klingeman, R.L. Beschta, P.D. Komar and J.B. Bradley (eds) *Gravel-Bed Rivers in the Environment*, 115–148, Highlands Ranch, CO: Water Resources Publications.
Best, J.L. and Ashworth, P.J. (1997) Scour in large braided rivers and the recognition of sequence stratigraphic boundaries, *Nature* 387, 275–277.
Colby, B.R. (1964) Scour and fill in sand-bed streams, *United States Geological Survey Professional Paper* 462-D.
Eaton, B.C. and Lapointe, M.F. (2001) Effects of large floods on sediment transport and reach morphology in the cobble-bed Sainte Marguerite River, *Geomorphology* 40, 291–309.
Haschenburger, J.K. (1999) A probability model of scour and fill depths in gravel-bed channels, *Water Resources Research* 35, 2,857–2,869.
Hassan, M.A. (1990) Scour, fill and burial depth of coarse material in gravel-bed streams, *Earth Surface Processes and Landforms* 15, 341–356.

Further reading

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

PETER ASHMORE

CYCLE OF EROSION

The Cycle of Erosion or 'The Geographical Cycle' was formulated in the latter years of the nineteenth century by W.M. Davis (e.g. Davis 1899). It was the first widely accepted modern theory of landscape evolution (see SLOPE, EVOLUTION). Davis regarded landscapes as evolving through a progressive sequence of stages, each of which exhibited similar landforms. In the Davisian model it was assumed that uplift takes place quickly. The land is then gradually worn down by the operation of geomorphological processes, without further complications being produced by tectonic movements. It was believed that slopes declined in steepness through time until an extensive flat region was produced close to BASE LEVEL, though locally hills called *monadnocks* might rise above it. This erosion surface was termed a *peneplain*. The reduction in the landscape creates a time sequence of landforms progressing through three stages: youth, maturity and old age.

Initially the Davisian model was postulated in the context of development under humid temperate ('normal') conditions, but it was then extended to other landscapes including arid (Davis 1905), glacial (Davis 1900), coastal (Johnson 1919), karst (Cvijić 1918) and periglacial landscapes (Peltier 1950).

Davis's model was immensely influential and dominated much of thinking in Anglo-Saxon geomorphology in the first half of the twentieth century, contributing to the development of DENUDATION CHRONOLOGY. Davis was a veritable 'Everest' among geomorphologists (Chorley *et al.* 1973). The model was largely deductive and theoretical and suffered from a rather vague understanding of surface processes, from a paucity of data on rates of operation of processes, from a neglect of climate change, and from assumptions he made about the rates and occurrence of uplift. However, it was elegant, simple and tied in with broad, evolutionary concerns in science at the time. Nonetheless, by the mid-1960s the concept was under attack (Chorley 1965).

The Davisian model was never universally accepted in Europe, where the views of W. Penck were more widely adopted. Penck's model involves more complex tectonic changes than that of Davis, and regards slopes as evolving in a different manner (slope replacement rather than slope decline) through time (Penck 1953). An alternative model of slope development by parallel retreat leading to *pediplanation* was put forward by L.C. King (e.g. King 1957). Thorn (1988) provides a comparative analysis of the models of Davis, Penck and King. Another evolutionary model of landscape evolution was produced by Büdel (1982), who developed the concept of ETCHING, ETCHPLAIN AND ETCHPLANATION.

References

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

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.

Chorley, R.J., Beckinsale, R.P. and Dunn, A.J. (1973) *The History of the Study of Landforms or the Development of Geomorphology. Volume 2. The Life and Work of William Morris Davis*, London: Methuen.

Cvijić, J. (1918) Hydrographie souterraine et evolution morphologique du karst, *Recueil des Travaux de L'Institut Géographie Alpine* (Grenoble), 6(4).

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

——(1900) Glacial erosion in France, Switzerland and Norway, *Proceedings of the Boston Society of Natural History* 29, 273–322.

——(1905) The Geographical Cycle in an arid climate, *Journal of Geology* 13, 381–407.

Johnson, D.W. (1919) *Shore Processes and Shoreline Development*, New York: Prentice Hall.

King, L.C. (1957) The uniformitarian nature of hillslopes, *Transactions of the Edinburgh Geological Society* 17, 81–102.

Peltier, L. (1950) The geographic cycle in periglacial regions as it is related to climatic geomorphology, *Annals of the Association of American Geographers* 40, 214–236.

Penck, W. (1953) *Morphological Analysis of Landforms*, London: Macmillan.

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

A.S. GOUDIE

CYCLIC TIME

A cycle is a period of time in which events happen in an orderly way. The order repeats itself in time so that there is a recurring series of changes. The term 'cyclic time' is unnecessary and illustrates the confusion that exists in the use of cyclic concepts.

Unfortunately a geomorphological cycle is often regarded as a sequence of *changes* from an initial state, through a series of stages to an ultimate state. In such models it is assumed that the changes taking place are such that the system has a different configuration when observed at different times, in other words, landforms have an observable history.

This emphasis meant that attention was paid to the sequence of changes and the 'stages of evolution' rather than the temporal lengths, frequencies and durations of the cycle and its events. This led to an emphasis on DENUDATION CHRONOLOGY (establishing and dating the stages of change) rather than a real understanding of geomorphological processes, rates of change and event statistics.

The interchangeable use of 'time' as the time during which changes take place, as the sequence of changes or the stage reached in the 'cycle' began with W.M.Davis (1899, 1905). In some passages he uses the correct dictionary sense. After describing the way a river advances through its long life and reduces an uplifted landmass to a PENEPLAIN he states, 'This lapse of time will be called a cycle in the life of a river.'

Unfortunately, he also described the geographical cycle as 'a complete sequence of landforms' but then qualified this as taking place from the uplift (an event) that produced the initial form through a sequence of form changes (responses to process events) to an ultimate form – a plain of low relief. In another passage Davis said that a geographical cycle 'may be divided into parts of unequal duration, each part of which will be characterized by the degree and variety of relief, and by the rate of change that has been accomplished since the start of the cycle'. Davis described how the successive forms of the cycle were dependent on three variable quantities: structure, process and time. He therefore makes it clear that 'time' is the amount of change from the initial form or its stage of development. In other passages the amount of change is 'a function of time' and time is again used as one of the trio of controls.

The period of time involved in a cycle has been poorly thought out. Davis (1899, 1905) estimated that the block mountains of Utah would be peneplained in 20–200 Ma. Wooldridge (personal, communication 1960) estimated up to 100 Ma but stated that the Mio-Pliocene peneplain had been produced in less than 20 Ma (Wooldridge and Linton 1955). Schumm and Lichty (1965) thought in terms of 10^6 years and Schumm (1963) pointed out that the time period to base levelling would be greatly extended by isostatic, erosional rebound. The general conclusion is that denudation cycles involve time spans of geological duration for their completion and recurrence by further uplift.

It is now known that the controls of Earth systems, such as structure, climate and base level, do not remain stable for such long periods of time and that it is preferable to establish the time periods for the frequency of landform creation events, the relaxation times and the landform survival times for the relevant system specifications. Some geomorphologists would argue (Schumm and Lichty 1965) that time can be divided into cyclic, graded and steady time periods. A more recent view (Graf 1977; Brunsden and Thornes 1979; Brunsden 1990) would suggest that the period 'cycle' be dropped in favour of system-based terms.

The name of a cycle (cyclic time?) is taken from the subject matter of the changes involved. General examples are a geographical cycle, a geomorphic cycle, an erosion cycle, a cycle of topographic development, a cycle of denudation, a cycle of life (Davis 1899). More specific uses were the normal cycle (landscapes developed under humid temperate conditions), shoreline development, sedimentation, karst, slope evolution, underground drainage, hydrologic, climatic and cycles of all geomorphological processes regimes.

References

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

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

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

——(1905) The Geographical Cycle in an arid climate, *Journal of Geology* 13, 381–407.

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

Schumm, S.A. (1963) Disparity between present rates of denudation and orogeny, *US Geological Survey Professional Paper* 454, 13.

Schumm, S.A. and Lichty, R.W. (1965) Time, space and causality in geomorphology, *American Journal of Science* 263, 110–119.

Wooldridge, S.W. and Linton, D.L. (1955) *Structure, Surface and Drainage in South-East England*, London: George Philip.

DENYS BRUNSDEN

CYMATOGENY

A term introduced by L.C. King (1959) to describe crustal movements intermediate between EPEIROGENY and OROGENESIS. They involve a warping of the Earth's crusts over horizontal distances that range from tens to hundreds of kilometres, and with vertical movements up to thousands of metres. They involve, however, minimal rock deformation. It is thought that the uplift is caused by processes active within the Earth's mantle.

Reference

King, L.C. (1959) Denudation and tectonic relief in southeastern Australia, *Transactions of the Geological Society of South Africa* 62, 113–138.

A.S. GOUDIE

D

DAM

Dams have been used to secure water supplies, to control floods and to generate power for more than a thousand years. The earliest civilizations developed along rivers in arid and semi-arid areas, such as along the Nile, and it is here that the oldest dams were built about 5,000 years ago. Flows in the wet season were stored in reservoirs to supply water for large-scale irrigation agriculture during the dry season. Water security and food security were closely linked and maintained the social, economic and political stability of the developing civilizations.

Today, the flows on most rivers are controlled to some degree by dams (WCD 2000). There are more than 45,000 dams over 15 m high and the largest dams stand more than 200 m high! The first big dam was the 221 m high Hoover Dam on the Colorado River, constructed in 1935. Kariba dam on the Zambezi, closed in 1958, was the first large dam to be constructed in the tropics. Water stored in reservoirs exceeds that stored in natural lakes by more than three times. Major rivers such as the Colorado and Columbia in USA, the Volga in Europe, the Nile in Africa, the Parana in Latin America, and the Murray–Darling in Australia have been intensively developed. Hydro-electric power is a major driver of dam building. Only about 3 per cent of the world's total energy consumption is supplied by water power and some 75 per cent of the hydroelectric power potential of the world's rivers is still to be exploited.

The geomorphological significance of large dams includes reservoir-induced earthquakes that have occurred in a small proportion of cases but have dramatic impacts. Large dams and reservoirs can both increase the frequency of earthquakes in areas prone to seismic activity and cause earthquakes in areas thought to be geologically stable. The mechanism involves the extra water pressure created by the dam and reservoir within faults in underlying rocks. Gupta (1992) records seventy examples of reservoir-induced seismicity. In many cases, the strongest shocks, often exceeding 4 and occasionally 6 on the Richter scale, occurred shortly after the initial filling of the reservoir.

Much more common are the impacts of dams on the fundamental fluvial processes, the flow and sediment transport regimes. These process changes induce adjustments of the size and shape of the river channel, and the form of the floodplain. These changes of the flooding and sedimentation regimes together with the changes to the morphology of the river corridor impact upon plants and wildlife by changing the habitats available for biota.

All dams are designed to capture floodwaters (see FLOOD) and represent perhaps the greatest point-source of hydrological impact. On some rivers reduced flood magnitudes have been experienced for more than 1,000 km below the dam and below the Aswan Dam on the River Nile, the reduction in freshwater flows is seen in the increased salinity of waters offshore of the delta in the south-east Mediterranean Sea. The Colorado River, USA, is dammed along its length, as are its major tributaries, and less than 1 per cent of the virgin flow reaches the river mouth. On the Murray–Darling system in Australia, which is regulated by nine principal storage reservoirs, the natural flow pattern was reversed, with high flows being released from the dams to supply irrigation demands downstream.

The basic concept of flood storage is 'empty space', keeping a reservoir as empty as possible to store floodwaters when they arrive. Water-supply reservoirs need to be kept full to provide water for domestic, industrial or irrigation supplies during dry seasons and dry years. But even when a reservoir is full and spilling over the dam, the flood peak downstream will not be as high as that for the inflow because of temporary storage in the lake as levels rise above the crest level of the overflow weir. Commonly, the size of the mean annual flood below dams has been reduced by between 25 and 50 per cent.

Dams and reservoirs also trap the sediments transported by a river – in many cases permanently storing the entire sediment load supplied by the upstream drainage basin. As the relatively high-velocity, and turbulent, water of a river feeding the reservoir is transferred into the slow-flowing water within the lake the sediment is deposited. Part is deposited in the reservoir itself and part in the channel and valley-bottom upstream, as a result of the backwater effects from the reservoir reducing velocities of river and floodplain flows. The coarser sediments settle out to form a delta. The finer particles, especially the clays, are distributed further out into the lake. Average annual rates of reservoir storage loss are usually less than 0.5 per cent per year but exceptional rates of more than 2 per cent per year have been reported from regions with high SEDIMENT LOAD AND YIELDS. One extreme case is the Heosonghi Reservoir on the Huang Ho, China that lost nearly 20 per cent of its storage capacity within three years of completion.

Flows released from dams or passing the spillway during floods are known as 'clearwater' releases because they are more or less sediment free. However, sometimes the water can appear turbid, not because of suspended sediments, but because of high concentrations of plankton when water is released from the lake surface during summer. This is caused by phytoplankton – algae and diatoms – which can reach high concentrations in relatively warm, surface layers of reservoirs having long retention times. Turbid releases may also be caused by the discharge of deep water during the autumn when stratified lakes mix – the 'overturn'. Such discharges can contain high concentrations of iron, manganese and hydrogen sulphide, giving a bad egg smell. However, in both cases, the quality of the water discharged from a reservoir can be controlled by the selective release of water from different depths within the lake. Occasionally, sediments are deliberately flushed from reservoirs by opening deep valves in the dam, to reduce the rate of storage loss. An example of this operation is the management of the Verbois, Chancy–Pougny and Genissiat reservoirs on the River Rhone in France. During these rare events, suspended sediment concentrations can exceed 1 g l^{-1} but such sudden surges of sediment-laden water can cause problems for water quality downstream.

Clearwater releases and the regulated flow regime below dams induce changes of channel morphology. The size and shape of natural river channels are in regime with the flows and sediment loads. Below dams two general types of change in regime can occur, although in detail there are many variations on these (Brandt 2000). The first type of channel change occurs where the dominant change of fluvial process is the reduction in sediment load. The clearwater releases of sediment-free water from reservoirs into channels with alluvial bed and banks can cause rapid erosion, or degradation, that may extend for many kilometres downstream. Typically bed degradation deepens the channel and the banks may also be undermined and sand and gravel bars eroded. An increase in the size of the sediments on the channel bed, which becomes armoured by the selective removal of the finer particles, and the reduction in channel slope as a result of bed incision may limit the amount of bed erosion. The result is a channel of increased cross-sectional area. Reports of degradation rates of more than 100 mm per year over channel lengths of more than 100 km are not uncommon. Rates decline over time until a new 'regime' condition is reached.

The second type of channel response is to the regulated flows, especially the lower flood levels. This induces a reduction of channel capacity most commonly observed as a reduction of channel width. Flow regulation reduces the capacity of a river to transport sediments supplied by sources downstream from the dam. These sources include tributary catchments and any degrading reaches and the dam and reservoir site during construction. Coarse sediments will be deposited on the channel bed but sediments will also accumulate as bars and benches along the channel margin, sometimes creating a new floodplain. The former floodplain is then converted into a river terrace (see TERRACE, RIVER).

The rate of channel narrowing is highly variable but can be particularly rapid in two situations. First, channel change is often rapid along

regulated rivers in semi-arid areas where wide, braided rivers are converted into single channels. In these cases, the growth of vegetation such as willows and poplars, sometimes accelerated by the maintenance of higher baseflows than in the natural river, can result in dramatic reductions of channel width (Merritt and Cooper 2000). The second situation is downstream from tributaries that produce high sediment delivery to the regulated channel. Sometimes, the reduced flood levels within the regulated river can accelerate erosion within the tributary increasing sediment supplies until the tributary has reached regime (Germanovski and Ritter 1988).

Each river comprises a sequence of channel reaches, each having a different channel form reflecting the history of the reach over Quaternary, historical and recent timescales. Channel change involves the movement – erosion, transport and deposition – of large volumes of sediments into and through this series of channel reaches over periods of time ranging from years to centuries. Volumes of up to 1 million cubic metres in a one-kilometre reach are not uncommon. In many cases individual reaches of river channel will show a COMPLEX RESPONSE to impoundment (Sherrard and Erskine 1991; Church 1995). This involves alternating phases of degradation and aggradation as the river network, the main channel and its tributaries, continue to adjust to the regulated flow regime by moving sediment through the sequence of reaches until a new 'regime' channel form is established.

Along rivers that have low sediment loads and stable, cohesive bank materials, adjustments of channel form may be very slow. In extreme cases the timescale for channel change to establish a new 'regime' channel may extend to hundreds of years. In these cases, the existing channel form will accommodate the regulated flows and evidence of upstream impoundment may be limited to local sediment accumulation in pools and backwaters, the growth of moss on large stones, and the marginal growth of emergent aquatic plants. An extreme flood may be required to initiate major channel changes in these reaches.

Geomorphology provides a physical template for river, riparian and floodplain ecology (Petts 2000) (see PHYSICAL INTEGRITY OF RIVERS). Variable river flows and sediment loads, and dynamic channels that change position by the processes of deposition and erosion, creating new floodplain patches and eroding others, sustain a diverse and highly productive riverine ecosystem. The channel pattern (see CHANNEL, ALLUVIAL) determines the range of habitat types found along any river but the frequency of erosion and deposition determine the level of disturbance that rejuvenates ecological successions.

Dams reduce the physical dynamism of the downstream riverine ecosystem, simplifying the physical habitat, and reducing both biological diversity and productivity (Ward and Stanford 1995). Advances in the application of geomorphological knowledge to the operational management of regulated rivers through the development of instream flow models (Petts and Maddock 1994) seek to sustain the ecological integrity of rivers below dams. Such models determine three levels of flows that need to be sustained along a regulated river to maintain the physical and, therefore, ecological dynamism of the river corridor. These flows are the floodplain maintenance flow, the channel maintenance flow (usually the BANKFULL DISCHARGE), and flushing flows to prevent the siltation of the channel bed and to prevent vegetation encroachment into the channel.

References

Brandt, S.A. (2000) Classification of geomorphological effects downstream of dams, *Catena* 40, 375–401.

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

Germanovski, D. and Ritter, D.F. (1988) Tributary response to local base level lowering below a dam, *Regulated Rivers* 2, 11–24.

Gupta, H. (1992) *Reservoir-Induced Earthquakes*, Amsterdam: Elsevier.

Merritt, D.M. and Cooper, D.J. (2000) Riparian vegetation and channel change in response to river regulation: a comparative study of regulated and unregulated streams in the Green River Basin, USA, *Regulated Rivers* 16, 543–564.

Petts, G.E. (2000) A perspective on the abiotic processes sustaining the ecological integrity of running waters, *Hydrobiologia* 422/423, 15–27.

Petts, G.E. and Maddock, I. (1994) Flow allocation for in-river needs, in P. Calow and G.E. Petts (eds) *The Rivers Handbook*, 2, 289–307.

Sherrard, J.J. and Erskine, W.D. (1991) Complex response of a sand-bed stream to upstream impoundment, *Regulated Rivers* 6, 53–70.

Ward, J.V. and Stanford, J.A. (1995) The Serial Discontinuity Concept: extending the model to floodplain rivers, *Regulated Rivers* 10, 159–168.

WCD (2000) *Dams and Development*, London: World Commission on Dams (WCD) and Earthscan.

Further reading

Beyer, P.J. (ed.) (2004) Dams and geomorphology, Binghamton Symposium Special Issue, *Geomorphology*, in press.

Petts, G.E. (1984) *Impounded Rivers*, Chichester: Wiley.

——(1994) Large-scale river regulation, in C.N. Roberts (ed.) *The Changing Global Environment*, 262–283, Oxford: Blackwell.

Williams, G.P. and Wolman, M.G. (1984) Downstream effects of dams on alluvial rivers, *US. Geological Survey Professional Paper* 1,286.

The journal *River Research and Applications* (until 2002 *Regulated Rivers*) has a focus on rivers below dams.

GEOFFREY PETTS

DAMBO

A headwater valley in areas of low relief, particularly in the seasonal tropics, that is channelless and in humid areas may contain swamps. Dambos are also known as *vleis* in southern Africa, *matoro* in Zimbabwe, *baixas* in Amazonia, *bolis* in Sierra Leone, *mbuga* in East Africa and *fadama* in northern Nigeria. German geomorphologists (e.g. Büdel 1982) have called them 'Spülmulden' or *wash depressions.* True dambos tend to be restricted to climates with present-day rainfalls between 600 and 1,500 mm, but the *bolis* of Sierra Leone are found where annual rainfall approaches 2,500 mm. They are also probably best developed on ancient planation surfaces. They occur on a wide range of rock types from unconsolidated Kalahari Sand through to shales, quartzites, schists, gneisses and granites (Thomas and Goudie 1985, Plate 31).

Their hydrology has been described by Bullock (1992), and they are a major source of water supply in rural areas in countries like Zimbabwe. Many of them are now being exploited for agricultural reasons and are suffering degradation, including gullying, as a consequence. Indeed, dambo is a Bantu word meaning 'meadow grazing,' for they are often grass covered and have no true woodland vegetation (Mäckel 1974).

Dambos tend to have low gradients (usually less than 2°). They receive their water either from direct precipitation onto the dambo or by subsurface flow from the surrounding high ground. With regard to the processes that lead to their formation, two main schools of thought exist (Boast 1990). The fluvial school envisages dambos as the simple extensions of the channelled drainage network. Rivers erode their head valleys which may subsequently be infilled by slope colluviation and by channel alluviation. Sheet-wash processes under seasonal rainfall regimes may be especially important. The other school of thought advocates differential chemical and biochemical corrosion or sapping rather than mechanical erosion as the main process. It sees dambo morphology as breaking 'too many fluvial rules' to be explicable in simple fluvial terms. That fluvial processes have operated in some dambos is made clear by the stratigraphy of their floors, which can reveal old alluvial fills. It is evident in many parts of central Africa that the balance between colluviation and alluviation has varied repeatedly in response to climatic changes. However, the two schools of thought are not necessarily mutually exclusive and Thomas (1994: 279) believes that 'Opposition between sapping (or etching) processes and sedimentation in dambos is misplaced.'

Plate 31 A broad, flat-floored, grassy dambo in west central Zambia

References

Boast, R. (1990) Dambos: a review, *Progress in Physical Geography* 14, 153–177.

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

Bullock, A. (1992) Dambo hydrology in southern Africa – review and assessment, *Journal of Hydrology* 134, 373–396.

Mäckel, R. (1974) Dambos: a study of morphodynamic activity on plateau regions of Zambia, *Catena* 1, 327–365.

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

Thomas, M.F. and Goudie, A.S. (eds) (1985) Dambos: small channelless valleys in the tropics, *Zeitschrift für Geomorphologie, Supplementband* 52.

A.S. GOUDIE

DATING METHODS

Stratigraphic relationships between landforms or within depositional sequences provide the most common, and simplest, means of deducing the age. Other than under exceptional circumstances, younger landscape features or sediments overlie older ones. However, this approach does not enable rates of processes to be deduced, nor give any idea of the relative or absolute timing of events. A number of dating methods exist based on chemical and biological changes that occur through time. The formation of CHEMICAL WEATHERING rinds and DESERT VARNISH on exposed rock are examples of the former, while amino-acid racemization and LICHENOMETRY are examples of the latter. These are all relative dating methods (i.e. indicating that one landform is approximately twice as old, or three times as old, as another). Another class of dating methods is that based on the correlation of events. For example, periodically the Earth's magnetic field is reversed, with the positions of the north and south magnetic poles switching. The last time that such a reversal occurred was 780,000 years ago. This event is recorded in a number of sedimentary and volcanic records and provides a synchronous marker across the globe, thus allowing one site to be correlated to another. In order to define the numerical age of this event (i.e. the age expressed as the number of years before present) a different class of dating methods are required – absolute age methods.

The discovery of radioactivity at the end of the nineteenth century provided the foundation for a suite of absolute dating techniques collectively known as radioisotopic methods. These all rely upon the fact that the rate at which a radioactive isotope of an element undergoes decay to produce another isotope (known as the daughter product) is constant, unaffected by any external controls such as temperature or pressure.

Radiocarbon dating was the first radioisotopic dating method to be widely applied starting in the 1950s. Carbon occurs as three isotopes, ^{12}C, ^{13}C and ^{14}C. The first two are stable isotopes, while the latter is radioactive, but all react chemically in identical ways. Radiocarbon (^{14}C) is generated in the upper atmosphere by the interaction of high energy cosmic rays with nitrogen atoms. The ^{14}C generated in this way is rapidly oxidized to form carbon dioxide which enters the carbon cycle. Radiocarbon has a half-life (the time taken for half of the atoms of ^{14}C within a sample to undergo radioactive decay) of $5{,}730 \pm 40$ years, and the concentration of ^{14}C in the atmosphere is a balance between the rate of production and the rate of decay. All living things exchange carbon with some part of the carbon cycle, and thus contain ^{14}C. After death this exchange ceases. The ^{14}C continues to decay according to its half-life, but it is no longer replaced by exchange with any part of the carbon cycle. Measurement of the ^{14}C remaining in a sample allows calculation of the period of time since death. Radiocarbon dating is most appropriate for organic materials, but can also be applied to some carbonates. The method assumes that the concentration of ^{14}C in the various reservoirs of the carbon cycle has remained constant through time. Measurement of the ^{14}C activity of tree rings of known age for the last 11,000 years shows this not to be the case, but these results allow ^{14}C ages to be calibrated to calendar years (Aitken 1990: 98). Between 11,000 years and ~40,000 years, the limit of the method, the ^{14}C calibration is less well known and the uncertainties on the ages larger.

In addition to ^{14}C, a wide variety of other isotopes (^{10}Be, ^{26}Al, ^{36}Cl) are generated both in the atmosphere and at the surface of the Earth by the interaction of cosmic rays. A suite of dating methods based on these cosmogenic isotopes have recently been developed (see COSMOGENIC DATING).

Other radioisotopic methods rely upon the very long half-lives of certain isotopes. Uranium occurs naturally as several isotopes (^{234}U, ^{235}U, ^{238}U). ^{238}U has a half-life of 4.47×10^9 years, comparable with the age of the Earth, and thus a significant quantity persists in the natural environment. Unlike ^{14}C, whose daughter product (^{14}N) is stable, the decay of ^{238}U produces ^{234}Th, which is itself radioactive. This in turn decays to produce ^{234}Pa, which decays to produce ^{234}U, ^{230}Th, ^{226}Ra and so on, producing a decay series until a stable isotope, ^{206}Pb, is produced (Figure 32). Over time the concentration of the different isotopes within the decay chain will alter until a state is reached where the number of decays per unit time from each isotope is identical – this state is termed secular equilibrium. The different chemical characteristics of the elements within the decay series provide a number of radioisotopic dating methods. For instance, when calcite is deposited in KARST environments, trace quantities of uranium are also deposited.

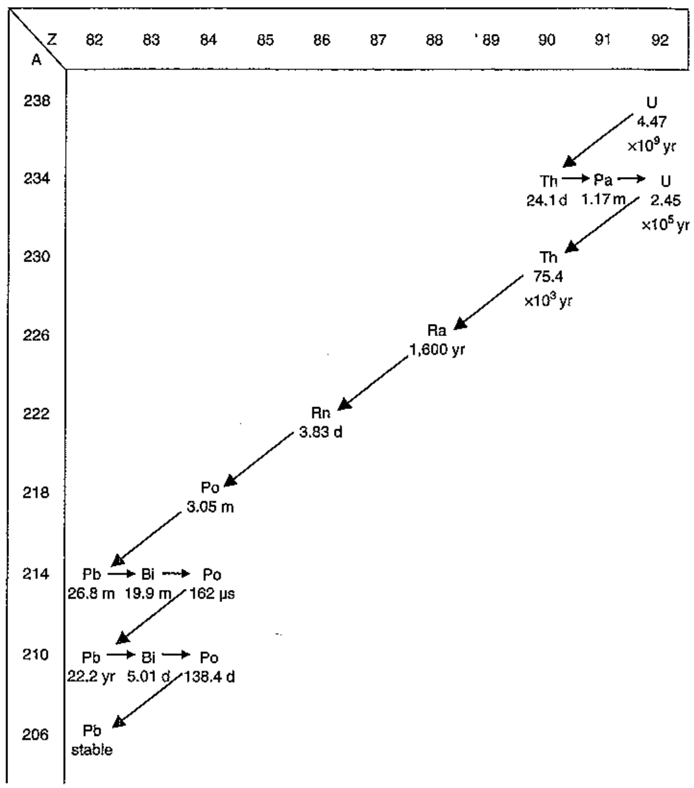

Figure 32 Decay series for ^{238}U. The half-life of each isotope is shown under the isotope. A is the atomic mass, while Z is atomic number

However, little or no thorium is deposited because it is relatively insoluble. Thus within the calcite, uranium will occur but without its thorium daughter products – it is said to be daughter-deficient. Over time, as the uranium undergoes decay, the concentration of thorium will increase. At the time of deposition, the ^{230}Th/^{234}U ratio will be zero, and will increase in a predictable manner allowing the age of formation of the calcite to be determined. This process can be used to date the precipitation of calcites over the last 350,000 years. As well as calcite in karst environments, another excellent target for Th/U dating is coral (Muhs 2002). Another part of the uranium decay series ^{210}Pb has a much shorter half-life (22 years) and can be used to provide ages over the last 100 years. In this case, the method is most commonly applied to lake sediments and nearshore marine basins (Appleby and Oldfield 1992). The method relies on the fact that one of the isotopes within the ^{238}U decay series, ^{222}Rn (radon), is a gas. This escapes to the atmosphere where it will undergo decay via a series of short-lived daughter products to produce ^{210}Pb. This falls from the atmosphere, producing a near constant supply to the surface of lakes and the nearshore. This ^{210}Pb is incorporated into the sediment accumulating under the water body, but none of its parent isotopes are present – thus there is a daughter excess.

Like uranium, potassium has an isotope with a long half-life (1.25×10^9 years). A small, but

nificant, proportion (0.01167 per cent) of all potassium is the radioactive isotope ^{40}K. This forms the basis for the techniques of potassium–argon (K–Ar) and argon–argon (Ar–Ar) dating of volcanic rocks (see VOLCANO). ^{40}K undergoes radioactive decay to produce either ^{40}Ca or ^{40}Ar. Argon is an inert gas, and while magma is molten any ^{40}Ar produced will be driven off, eventually making its way into the atmosphere (where it constitutes ~1 per cent by volume). Once crystallization occurs at the time of eruption, argon is unable to escape and begins to accumulate within the minerals crystal structure. Thus the ratio of the parent isotope (^{40}K) to the daughter product (^{40}Ar) provides a means of dating the volcanic eruption – this is the K–Ar method. The ratio of the parent and daughter isotopes can be measured more precisely by irradiating the sample of volcanic tephra or lava in a neutron beam in a nuclear reactor. This causes a proportion of the potassium to transform to ^{39}Ar, an isotope not found in nature. The age of the sample can then be found by measuring the ratio of two argon isotopes, ^{39}Ar (which is now a measure of the potassium concentration) and ^{40}Ar. Measuring this isotopic ratio is a more precise analytical process than measuring potassium and argon separately. Equally importantly, both argon isotopes are measured on the same subsample, thus allowing samples as small as single tephra crystals to be dated. Using the Ar–Ar method, ages as recent as a few thousands of years can be obtained (e.g. Renne *et al.* 1997, Figure 33).

An alternative approach to dating is not to measure the concentration of radioactive isotopes directly, but instead to look at the effect that the radioactivity has on materials in the natural environment – these are radiogenic methods. One such method is fission track dating. The most common way in which uranium decays is by the emission of an alpha particle (consisting of two neutrons and two protons). However, ^{238}U may also undergo fission, whereby the nucleus (consisting of 92 protons and 146 neutrons) splits into two new nuclei of almost equal masses. A significant amount of energy is released at the same time, and the two nuclei (the fission fragments) recoil away from each other. This leads to ionization of the crystal along these tracks – this damage can be made visible by etching the crystal surface using acids, and the number of fission tracks counted. The method is most commonly applied to volcanic rocks, including far-travelled tephra, and dates the formation of the crystals. Zircons have the advantage of high uranium concentrations (typically between 10 and 1,000 ppm) meaning that the number of tracks produced in a given time will be high. Glass has a much lower uranium concentration (~1 ppm) but is the most abundant component of tephra, and it too can be used for fission track dating providing that a method such as Isothermal Plateau Fission Track Dating (ITPFT) is used which compensates for the ability of glass to naturally anneal fission tracks (Westgate 1989).

Luminescence techniques are also based on the effects of radioactive decay. Alpha, beta and gamma radiation, resulting from the decay of various radioactive elements in the Earth's crust, is ubiquitous. When this radiation is absorbed by commonly occurring minerals such as quartz and feldspar, the energy from the radiation may be used to trap electrons at excited sites within the crystal. In effect, the mineral grains act as dosemeters, integrating the total amount of radioactivity that they are exposed to. In the laboratory, these mineral grains can be stimulated, allowing the trapped electrons to release their stored energy. The energy is released as light emitted from the quartz or feldspar grains – it is this light that is called luminescence. If the mineral grains are stimulated by heating (typically up to 500 °C) then this is termed thermoluminescence (TL). For geological materials it is normally more appropriate to stimulate them using light of a fixed wavelength (e.g. 532 nm from a $NdYVO_4$ laser) in which case optically stimulated luminescence (OSL) is observed. The luminescence signal is light sensitive, and exposure to natural daylight reduces the luminescence signal to a low level. Many subaerial transport processes will entail exposure of mineral grains to daylight (e.g. AEOLIAN PROCESSES) and the sediments deposited by these processes (e.g. DUNE, AEOLIAN; LOESS) are ideally suited to luminescence dating (Stokes 1999). Upon burial the continued exposure to radiation from the natural environment causes the trapped electron population to increase with time. The OSL signal is reset by exposure to daylight more completely than the TL signal, and hence the use of OSL has allowed more precise ages to be obtained and has allowed younger samples to be dated. In environments where the exposure to daylight at deposition can be assumed, events as recent as the last 50–100 years can be routinely dated.

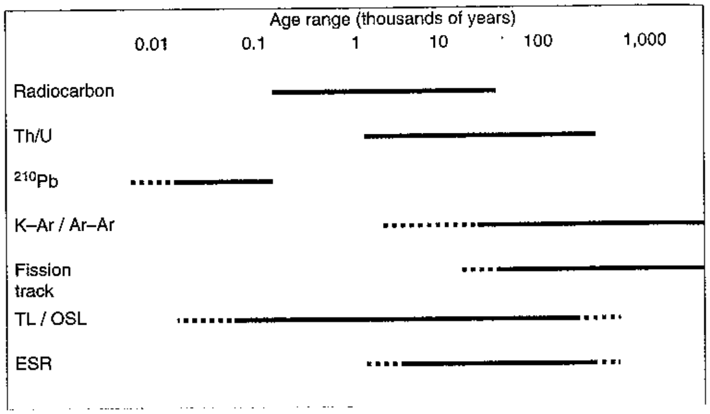

Figure 33 Age ranges over which various radioisotopic and radiogenic dating methods can be applied. The exact limits are often determined by the nature of the material being dated, and the dashed lines reflect this variation from one application to another

Electron spin resonance (ESR) dating is another technique based on measurement of the charge trapped in materials due to radiation from the environment. While TL and OSL are applicable to quartz and feldspar in sediments, ESR can be applied to stalagmites, tooth enamel, corals and sometimes bones.

References

Aitken, M.J. (1990) *Science-based Dating in Archaeology*, Harlow: Longman.
Appleby, P.G. and Oldfield, F. (1992) Application of Lead – 210 to sedimentation studies, in M. Ivanovich and R.S. Harmon (eds) *Uranium-series Disequilibrium, Applications to Earth, Marine and Environmental Sciences*, 2nd edition, 731–778, Oxford: Clarendon Press.
Muhs, D.R. (2002) Evidence for timing and duration of the last interglacial period from high-precision uranium-series ages of corals on tectonically stable coastlines, *Quaternary Research* 58, 36–40.
Renne, P.R., Sharp, W.D., Deino, A.L., Orsi, G. and Civetta, L. (1997) Ar-40/Ar-39 dating into the historical realm: calibration against Pliny the Younger, *Science* 277, 1,279–1,280.
Stokes, S. (1999) Luminescence dating applications in geomorphological research, *Geomorphology* 29, 153–171.
Westgate, J.A. (1989) Isothermal plateau fission-track ages of hydrated glass shards from silicic tephra beds, *Earth and Planetary Science Letters* 95, 226–234.

Further reading

Ivanovich, M. and Harmon, R.S. (eds) (1992) *Uranium-series Disequilibrium, Applications to Earth, Marine and Environmental Sciences*, 2nd edition, Oxford: Clarendon Press.
Noller, J.S., Sowers, J.M. and Lettis, W.R. (eds) (2000) *Quaternary Geochronology: Methods and Applications*, Washington, DC: American Geophysical Union.
Taylor, R.E. and Aitken, M.J. (1997) *Chronometric Dating in Archaeology*, New York: Plenum Press.
Wagner, G.A. (1998) *Age Determination of Young Rocks and Artifacts*, Berlin; New York: Springer.
Wintle, A.G. (1996) Archaeologically-relevant dating techniques for the next century, *Journal of Archaeological Science* 23, 123–138.

SEE ALSO: cosmogenic dating; dendrochronology; lichenometry

G.A.T. DULLER

DAYA

Small, silt-filled, closed solutional depressions found on limestone surfaces in some arid areas of the Middle East and North Africa. They are a type of PAN.

Reference

Mitchell, C.W. and Willmott, S.G. (1974) Dayas of the Moroccan Sahara and other arid regions, *Geographical Journal* 140, 441–453.

A.S. GOUDIE

DEBRIS FLOW

Debris flows are MASS MOVEMENT phenomena transitional between LANDSLIDES and sediment-laden water floods. They occur commonly in tectonically active regions subject to rapid uplift and erosion. Typically debris flows consist of churning, water-saturated mixtures of poorly sorted sediment and miscellaneous detritus, which rush down slopes and funnel into channels when they reach valley floors. Debris flows generally form abrupt surge fronts, attain peak speeds greater than 10 metres per second, and include up to 70 per cent solid particles by volume. As a consequence, debris flows can denude slopes, damage structures, drastically alter stream channels and endanger human life. Notable debris-flow disasters include those in Armero, Colombia, 1985, and Vargas state, Venezuela, 1999, each of which resulted in more than 20,000 fatalities.

Debris flows have some alternative names. For example, LAHAR is a commonly used Indonesian term for a debris flow that originates on a volcano, and mudflow describes a debris flow that consists predominantly of silt and clay. Such fine-grained flows are rare in SUBAERIAL settings but more common in submarine (see SUBMARINE LANDSLIDE GEOMORPHOLOGY) environments.

Most subaerial debris flows commence as rapid landslides triggered by intense rainfall or rapid snowmelt. A flow may originate from a single, discrete landslide source or from numerous, distributed sources from which debris issues and coalesces. Source areas generally slope more steeply than 25 degrees, but debris flows commonly scour bed and bank sediment from channels that slope as gently as about 8 degrees. On flatter slopes debris flows typically decelerate and form lateral LEVEES and lobate deposits that are very poorly sorted and readily distinguished from fluvial deposits. Many ALLUVIAL FANs in tectonically active regions are composed largely of debris-flow deposits.

Debris flows have a remarkable ability to flow quite fluidly, despite having grain concentrations comparable to those of static soil. The fluidity of debris flows results principally from a phenomenon called LIQUEFACTION, which occurs when pressure in the intergranular pore water rises to levels sufficient to support the weight of the overlying debris, thereby reducing friction at grain contacts. The reduced friction allows grains to move smoothly past one another, facilitating downslope flow. Liquefaction commences when debris flows begin to mobilize during landsliding of loosely packed soil or sediment, which contracts during shear deformation and transfers pressure to the intergranular pore water. Liquefaction persists in debris-flow bodies because silt and clay-sized sediment impedes pore-pressure dissipation, even if the fine sediment comprises just a few per cent of the debris-flow mass.

Effects of liquefaction are reduced or absent at the heads and lateral margins of debris-flow surges, where high concentrations of coarse debris accumulate. Debris-flow deposition occurs because coarse-grained marginal debris lacks high pore-water pressures and exerts strong frictional resistance to motion.

Further reading

Iverson, R.M. (1997) The physics of debris flows, *Reviews of Geophysics* 35, 245–296.
Iverson, R.M., Reid, M.E. and LaHusen, R.G. (1997) Debris-flow mobilization from landslides, *Annual Review of Earth and Planetary Sciences* 25, 85–138.
Johnson, A.M. (1984) Debris flow, in D. Brunsden and D.B. Prior (eds) *Slope Instability*, 257–361, Chichester: Wiley.
Takahashi, T. (1991) *Debris Flow*, Rotterdam: Balkema.

RICHARD M. IVERSON

DEBRIS TORRENT

Debris torrents are a regional phenomenon, extensively documented in the coastal Pacific north-west of the United States, British Columbia and south-east Alaska. A debris torrent is defined as 'a mass movement event that involves water-charged, predominantly coarse grained inorganic and organic material flowing rapidly down a steep, confined, pre-existing channel' (Van Dine 1985; Slaymaker 1988). This is a North American usage which contrasts with the European usage of the term torrent (*torrent* in French; *torrente* in Italian and *wildbach* in German). In Europe, torrent is descriptive of mountain stream morphology and not of a debris discharge event (Aulitzky 1980). Descroix and Gautier (2002) describe the appearance and disappearance of torrents (in the sense of a distinctive morphology) in the southern French Alps as a function of climate and land use changes.

Swanston (1974) and Hungr *et al.* (1984) have argued that the term 'debris torrent' is highly descriptive and well suited to the particular character of coarse-grained, channelized mass movement events of the Pacific maritime mountains. Slaymaker (1988) has argued that the case for debris torrents as a separate category is that they are a form of channelized debris flow which lack a fine-grained fraction, particularly clay, and have a relatively large organic debris content.

Debris torrents tend to occur in small drainage basins, from 0.1–10 km^2 (Mizuyama 1982); have steep channels, with an initiation zone greater than 25°, an erosion/transport zone (10–25°) and a depositional zone (5–12°); occur in high runoff intensity zones and require substantial amounts of organic and inorganic debris available for mobilization. Triggering mechanisms include storm and/or snowmelt runoff, water release from subglacial or lake storage, log jam bursts, rockfall, debris or snow avalanches from upslope or seismic shaking. The history of sediment accumulation in the channel is also critical (Bovis and Dagg 1987). Little cohesive material is present in debris torrents, a high proportion is gravel and boulders and wood and organic mulch is prominent. A frontal and lateral 'macrostructure' consists of framework supported boulders which are pushed forward by a turbulent slurry. The slurry is extruded through the macrostructure, effectively producing a two-phase flow.

References

Aulitzky, H. (1980) Preliminary two-fold classification of torrent, *Symposium Interpraevent* 4, 285–309.
Bovis, M.J. and Dagg, B. (1987) A model for debris accumulation and mobilization in steep mountain streams, *Hydrological Sciences Journal* 33, 589–605.
Descroix, L. and Gautier, E. (2002) Water erosion in the southern French Alps: climatic and human mechanisms, *Catena* 50, 53–85.
Hungr, O., Morgan, G.C. and Kellerhals, R. (1984) Quantitative analysis of debris torrent hazards, *Canadian Geotechnical Journal* 21, 663–677.
Mizuyama, T. (1982) Analysis of sediment yield and transport data for erosion control works, *International Association of Hydrological Sciences Publication* 137, 177–182.
Slaymaker, O. (1988) The distinctive attributes of debris torrents, *Hydrological Sciences Journal* 33, 567–573.
Swanston, D.N. (1974) Slope and stability problems associated with timber harvesting, *USDA Forest Service, Pacific Northwest, General Technical Report* PNW–21.
Van Dine, D.F. (1985) Debris flows and debris torrents in the southern Canadian Cordillera, *Canadian Geotechnical Journal* 22, 44–68.

SEE ALSO: debris flow; mass movement

OLAV SLAYMAKER

DECOLLEMENT

A fault surface marking where crustal deformation occurs in a parallel fashion, usually between an upper mechanically weak horizon, layer, or boundary, and a lower undeformed boundary. Decollements or decollement surfaces are formed by the upper rock series sliding over the lower during folding, and so is associated with overthrusting. They are typical between crystalline basement rock overlying sedimentary rock, often in thrust faulted regions such as the Alps, the Jura Mountains and the Zagros Mountains of Iran.

Further reading

Ramsay, J.G. and Huber, M.I. (1987) *The Techniques of Modern Structural Geology – Volume 2: Folds and Fractures*, London: Academic Press.

SEE ALSO: crustal deformation

STEVE WARD

DEEP-SEATED GRAVITATIONAL SLOPE DEFORMATION

Deep-seated gravitational slope deformations (DGSDs) are gravity-induced processes which evolve over a very long time interval and usually affect entire slopes, displacing rock volumes up to hundreds of millions of cubic metres over areas of several square kilometres with thicknesses of several tens of metres. The main feature of these processes is the probable absence of a continuous surface of rupture and the presence, at depth, of a zone where displacement takes place mostly through microfracturing of the rock mass (Radbruch-Hall 1978). Before the definition in literature of DGSDs, Terzaghi (1950: 84) contributed to this subject significantly by clarifying the difference between 'creep' and 'LANDSLIDE' with a statement that is applicable also to deep-seated phenomena:

> A landslide is an event which takes place within a short period of time as soon as the stress conditions for the failure of the ground located beneath the slope are satisfied. By contrast, creep is more or less a continuous process. A landslide represents the movement of a relatively small body of material with well-defined boundaries, whereas creep may involve the ground located beneath all the slopes in a whole region and no sharp boundary exists between stationary and moving material.

Thus the deformation phase may be naturally followed by a sliding phase within which shear planes are recognizable, though the evolution time of these processes is hard to predict and generally extremely long.

DGSDs, thus defined by Malgot (1977), have been documented almost everywhere in the world since the end of the 1960s and described by different authors with different terms, such as sackung, gravity faulting, depth creep of slopes, deep-reaching gravitational deformations, deep-seated creep deformations, gravitational block-type movements, gravitational spreading and gravitational creep (see MASS MOVEMENT). In spite of the variety of terms used, at present the terms most frequently used to identify the main DGSD types are *sackung* and *lateral spreading*.

Sackung

SACKUNG can be described as a sagging of a slope due to visco-plastic deformations taking place at depths which affect high and steep slopes made up of homogeneous, jointed or stratified rock masses showing brittle behaviour (Zischinsky 1969; Bisci *et al.* 1996). Typical morphological features are twin ridges, trenches, gulls and uphill facing scarps in the upper part of the slopes whereas the middle and lower parts of the slopes tend to assume a convex shape because of bulging and cambering. At the foot of the slope sub-horizontal joints can be found. The displacement mechanism, though, has not been well defined. It is thought that the rock mass behaviour at depth is different from that at the surface, owing to the high confining pressure acting all over the material. Two main displacement models have been defined. Most researchers (e.g. Mahr 1977) assume that at depth, in correspondence with the central portion of the slope, a high confining pressure does not allow the formation of well-defined surfaces of rupture, permitting only viscous deformations (non-shearing model). On the contrary, at the top and toe of the slope, where these pressures are lower, such surfaces might develop. Savage and Varnes (1987) assume instead that the zone subject to ductile deformation is indeed interrupted along a shear surface located at the base of the unstable rock mass (plastic failure model).

Lateral spreading

Lateral spreading consists of lateral expansions of rock masses occurring along shear or tensile fractures. Two main types of rock spreading, occurring in different geological situations, can be distinguished (Pasuto and Soldati 1996):

1. *Lateral spreading affecting brittle formations overlying ductile units*, generally due to the deformation of the underlying material. They are characterized by prevalently horizontal movements along tensile fractures or subvertical tectonic discontinuities. Trenches, gulls, grabens, karst-like depressions in the competent rocks and bulges in the clayey material are common features in this type of deformation. The overburden of the rock slabs is generally assumed as the cause of long-term displacements affecting the underlying formations which result in the squeezing out of the weaker rock types and rock block spreading due to tensile stresses. The process may be accelerated by water percolation through the fissures and consequent softening of the clay shales. Downcutting of valleys may then induce rotational slides and rock falls, together with block tilting and rotation which may prepare the way for block slides. The process may continue and cause progressive spreading and dismembering of the rock slab. The spreading may extend for several kilometres back from the edges of plateau.
2. *Lateral spreading in homogeneous rock masses* (usually brittle) without a recognized or well-defined basal shear surface or zone of visco-plastic flow. Typical morphological evidence is given by double ridges, uphill-facing scarps, ridge-top depressions and infilled troughs. This phenomenon has been recognized as prevalent in high mountain areas. The pre-existence of cracks in the rock mass and a high relief energy are considered as favouring factors but the mechanics of the deformation have not yet been well defined.

The evolution scenarios of sackung and lateral spread are different. The former may be considered as an initial stage of rotational– translational slides, with the tendency to evolve into rock or debris avalanches, i.e. processes which may induce high geomorphological risk situations. On the other hand, the latter may correspond to an early phase in the development of block slide-type phenomena, which are usually subject to a slow evolution of displacements.

References

Bisci, C., Dramis, F. and Sorriso-Valvo, M. (1996) Rock flow (sackung), in R. Dikau, D. Brunsden, L. Schrott and M.-L. Ibsen (eds) *Landslide Recognition: Identification, Movement and Causes*, 150–160, Chichester: Wiley.

Mahr, T. (1977) Deep-reaching gravitational deformations of high mountain slopes, *Bulletin International Association of Engineering Geologists* 16, 121–127.

Malgot, J. (1977) Deep-seated gravitational slope deformations in neovolcanic mountain ranges of Slovakia, *Bulletin International Association of Engineering Geologists* 16, 106–109.

Pasuto, A. and Soldati, M. (1996) Rock spreading, in R. Dikau, D. Brunsden, L. Schrott and M.-L. Ibsen (eds) *Landslide Recognition: Identification, Movement and Causes*, 122–136, Chichester: Wiley.

Radbruch-Hall, D.H. (1978) Gravitational creep on rock masses on slopes, in B. Voight (ed.) *Rockslides and Avalanches*, 607–675, Amsterdam: Elsevier.

Savage, W.Z. and Varnes, D.J. (1987) Mechanics of gravitational spreading of steep-sided ridges ('Sackung'), *Bulletin International Association of Engineering Geologists* 35, 31–36.

Terzaghi, K. (1950) Mechanism of landslides, in S. Paige (ed.) *Application of Geology to Engineering Practice*, 83–123, Washington, DC: Geological Society of America.

Zischinsky, Ü. (1969) Über Sackungen, *Rock Mechanics* 1(1), 30–52.

Further reading

Cruden, D.M. and Varnes, D.J. (1996) Landslides types and processes, in A.K. Turner and R.L. Schuster (eds) *Landslides: Investigation and Mitigation*, 36–75, Washington, DC: Transportation Research Board, National Academy of Sciences, Special Report 247.

MAURO SOLDATI

DEEP WEATHERING

Weathering studies have enjoyed a precarious role in geomorphology, at once central and yet often neglected. Rock decay due to chemical and biochemical processes mediates the rate of erosion and destruction of relief in almost all climates, and the dominance of quartz sand in clastic sediments demonstrates its effectiveness. Soil clays are products of these processes and are universally recognized without special comment. However, weathered materials frequently extend well below the classic soil profile to depths of tens of metres, and not infrequently to more than 100 m. The transition from surface soil to fresh rock is described as the REGOLITH or *weathering profile*. While there is no formal definition of what constitutes *deep weathering*, some authors use the term to describe 'exceptional' depths of rock decay (Taylor and Eggleton 2001), but this reflects experience from outside the humid tropics where weathering depths exceeding 30 m are common (Plate 32). A different approach refers to denudation being 'weathering limited' where altered materials are removed more or less instantaneously following breakdown (by chemical and mechanical processes), or 'erosion limited' where stores of non-cohesive, weathered material underlie the landsurface. In the latter case the products of weathering have remained *in situ* for an unspecified period, and it implies that during this time rates of weathering have exceeded rates of erosion. It is these circumstances that lead to the formation of deep weathering profiles, often over periods of 10^6–10^7 y. In the upper zones of many deep weathering profiles the rock has been largely reduced to a mixture of clays, Al and Fe oxides and quartz sand, through which traces of the rock structure can still be seen. This material is termed SAPROLITE.

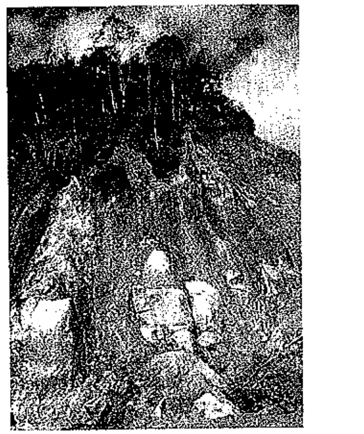

Plate 32 Deep weathering profile (>50 m) in granite with corestones in east Brazil

It is often stated that deep weathering is mainly associated with ancient landsurfaces of low relief and is the product of a humid tropical or subtropical environment. This reasoning is commonly applied to occurrences of deep weathering found in high latitudes, which are explained as relicts of a formerly extensive mantle of weathered rock formed at the end of the Mesozoic or in the early Cenozoic, when warm moist conditions prevailed to perhaps 60°N. In support of this view, extensive deep weathering of the Scandinavian shield rocks is found below Cretaceous sediments in South Sweden (Lidmar Bergström 1989), and 5–10 m of advanced alteration is found between Palaeogene lava flows in Northern Ireland (Smith and McAlister 1995). Deep saprolites are widely encountered throughout Western Australia, to depths of 100 m in places, and oxygen isotope and other methods have indicated ages from Permean to Miocene, when the Australian plate was far south of its present position and never in tropical latitudes (Bird and Chivas 1988). This led Taylor *et al.* (1992) to argue that time rather than climate might be the main determinant of advanced rock decay to great depths. However, deep saprolites exhibiting advanced weathering are found in Neogene terrain in the humid tropics, and have been cited from Borneo and New Guinea (Thomas 1994; Löffler 1977). Extensive planation is not recorded in these areas, so the profiles indicate high rates of weathering combined with low rates of erosion in a landscape of moderate relief in an equatorial climate under rainforest. In contrast, many deep weathering occurrences in high latitudes present features indicative of incipient rather than advanced decay. These materials are sandy, with a low clay content (typically 2–7 per cent), and are described as *arènes* (French) or GRUS (German). Occurrences of grus are found worldwide in temperate climates, and a similar material may be found at depth beneath clayey saprolites in the tropics. Grus depths are usually <15 m and commonly 3–6 m, but are not confined to landscapes of low relief (Migoń and Lidmar Bergström 2001). When all types and degrees of rock alteration are grouped together, deep weathering is found to be very widespread. It is comparatively rare in hot and cold deserts, and in areas of recent or active tectonics. Most of the regolith mantle has also been removed where there has been severe Pleistocene glacial scour. But deep profiles have been found in north-east Scotland (Hall 1985) and northern Scandinavia, where ice sheets were either cold based and non-erosive or had extended on to low ground.

The formation of deep weathering profiles poses difficult problems. For example, weathering processes are advanced by renewal of ground water and removal of minerals in solution, and will be inhibited by rising concentrations of solutes. Very deep profiles beneath ancient plateaux must, by this reasoning, require very long periods to form and need some means to export minerals in solution. Low solute concentrations in tropical rivers draining weathered landscapes are often cited in support of low weathering rates in these landscapes. The formation of a thick layer of saprolite is, therefore, considered by many to be a self-limiting system experiencing negative feedback. However, we know little about either the deep circulation of water or the potential for long distance migration of ions by diffusion processes. Arguments have been advanced in favour of hydrothermal processes being responsible for much deep rock decay, especially in granites. But many analyses have adduced oxygen isotope evidence for low temperature alteration (70 °C) as at St Austell, south-west England (Sheppard 1977), and hydrothermal mineralization is usually very restricted in extent (Ollier 1983). It is necessary to recognize the importance of interactions between meteoric water penetrating from the Earth's surface and juvenile waters generated by magmatic processes. Both are part of the global water cycle, and rock decay is ultimately a process of adjustment of mineral species to atmospheric conditions at the Earth's surface.

The existence of an extensive mantle of residual weathering products has great significance for engineers, as well as for geomorphologists and pedologists. But the nature of the weathered material is equally important. Grus behaves very differently from a clay-rich saprolite, for example. The transition from fresh rock, upward through the weathered rock towards the surface soil can be complex, but models have been developed to describe the *weathering profile*, as distinct from descriptions of soil profiles (Figures 34, 35). At the base of the profile is the WEATHERING FRONT, often described as the *basal weathering surface* because of the commonly observed, abrupt transition from sound rock to a disaggregated and altered 'saprock'. Very little chemical change is required to cause expansion of rock minerals by hydration and partial hydrolysis, leading to a disruption of the rock fabric. The most commonly described weathering profiles (Figure 34) are based on examples in jointed granites, and similar features are found in basaltic lavas and in feldspathic sandstones. But in banded and foliated metamorphic rock, such as schists, profile subdivisions may be indistinct.

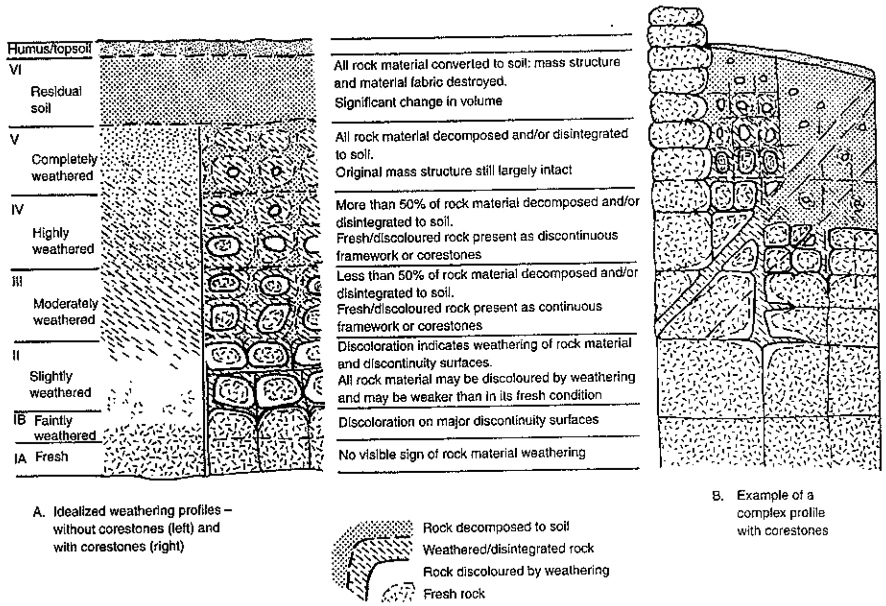

Figure 34 Characteristic weathering profiles with commonly used weathering grades shown in far-left column. Compiled by the author for Fookes (1997)

The 'granite model', first formally described from Hong Kong (Ruxton and Berry 1957), has been refined for the use of engineers (Fookes 1997); other models have been developed to describe mineralogic changes or the occurrence of specific weathering zones, including *laterite* (Figure 35). Chemical and mineralogical changes down profile are important in mineral prospecting, and the nature of the clays predicts engineering behaviour. The understanding of complete regolith profiles can be difficult due to problems of sampling, complexities of rock structure, and the mineral transformations caused by changing hydrologic conditions over long periods. But the issue is important if partly eroded (truncated) profiles, often found in the field, are to be correctly described and understood. The properties of soils in areas of deeply weathered rock, are strongly influenced by the degree of pre-weathering, which limits the availability of cations for plant growth. In many parts of the tropics, several generations of soils may have been formed, lost by erosion and re-formed within deeply weathered parent materials (Ollier 1959).

In TROPICAL GEOMORPHOLOGY, the role of the weathered mantle in determining landscape forms has been widely discussed (Thomas 1994). The balance between the rate of weathering and the rate of erosion is central to questions about the degree of alteration of near-surface weathering products on the one hand and the exposure of fresh rock forms on the other. Estimated rates of weathering on silicate rocks range from 2–$50\,\mathrm{m\,Ma^{-1}}$ ($\mathrm{mm\,ka^{-1}}$). Although surface erosion rates may exceed the highest value by two orders of magnitude, many forested slopes of moderate inclination in the tropics erode at rates less than $5\,\mathrm{mm\,ka^{-1}}$. But data are sometimes contradictory and it is difficult to generalize. Circumstantial evidence for low rates of erosion in undulating, forested terrain comes from the partial conformity of weathering zones with present-day relief, which often exhibits multiconvex weathered compartments (*demi-oranges*,

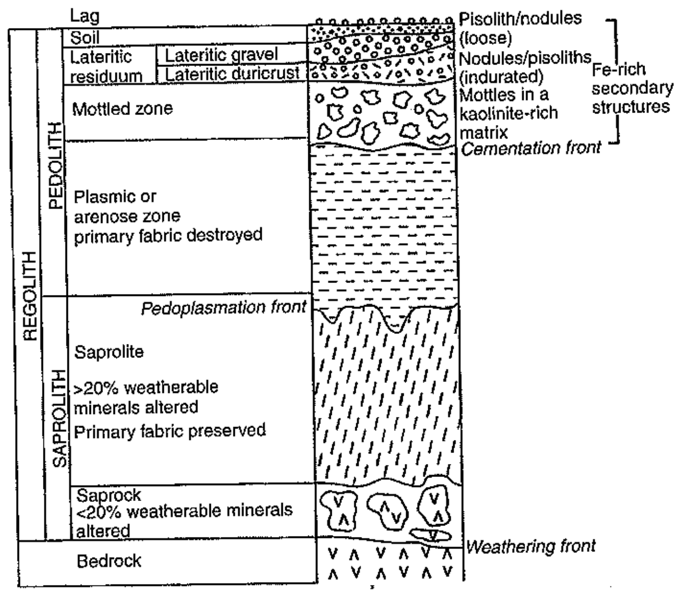

Figure 35 Scheme for regolith terminology in a profile with laterite. From Eggleton (2001)

French; meias laranjas, Portuguese). In the more arid areas of Africa and Australia, weathering profiles have been widely truncated, leaving mesa-shaped tabular hills capped by FERRICRETE or SILCRETE DURICRUSTS or landscapes with shallow, sandy regoliths and frequent outcrops (boulders, TORS, INSELBERGS). In central Australia, reference is made to 'the weathered landsurface' (Mabbutt 1965), and to the varied results of partial stripping of the regolith. Such landscapes have been described as varieties of 'etched plain' (see ETCHING, ETCHPLAIN AND ETCHPLANATION). A wider interpretation views these characteristics as the result of a 'cratonic regime' (Fairbridge and Finkl 1980) involving landscape stability and advanced weathering lasting perhaps 10^7–10^8 y, alternating with periods of erosion and regolith with a duration of 10^5–10^6 y.

In detail, the patterns of deep weathering can be shown to follow petrographic variations and structural weaknesses within rock masses. Ferromagnesian minerals and plagioclase feldspars decay more rapidly than orthoclase and mica in granites, and adjacent plutons containing different mineral suites often show contrasts in weathering. Potassium-rich intrusive rocks and silicified metamorphic gneisses, in particular, resist chemical attack. Intersecting joint patterns often outline basins of deeper rock decomposition. Deep erosion into ancient granite plutons exposes massive compartments under compressive stress that resist weathering although subject to spalling, while younger, higher level intrusives are usually subdivided along many joint directions. Where geology is uniform, patterns of weathering often respond to the relief, deeper weathering being found beneath convex summits in forested environments. While this can result from dissection into an extensive, deep saprolite mantle, the better drained conditions beneath upper slopes contribute to more rapid decay. Most perennial rivers flow in bedrock channels, but some channels in plateau landscapes alternate between anastomosing reaches containing rapids marking exposed fresh rock and meandering channels where the river flows above saprolite.

Deeply weathered landscapes were formally very extensive in Europe (and elsewhere). So-called 'lateritic' weathering covers were partially stripped from the Hercynian massifs of Europe during the Cenozoic, and are found today in the deposits of the Aquitaine, Paris and many other sedimentary basins. These were described by Millot (1970) as 'siderolithic facies', and elsewhere as 'laterite derived facies' (Goldberry 1979) and 'red beds'. During the Neogene a renewal of the regolith cover occurred patchily in the broken relief, resulting from Alpine tectonics. But the short duration of this period and the cooler climates of higher latitudes resulted in thinner, poorly differentiated sandy 'grus'. In the tropics, the breakup of Gondwanaland and drifting apart of the continents during the last 100 Ma also led to deep dissection and the infilling of downwarped and faulted sedimentary basins with the detritus of Mesozoic weathering. These are known as the Continental Terminal in west Africa and the Barreiras Formation in South America. Climatic vicissitudes have involved aridification of many tropical areas after the mid-Miocene, halting the advance of the weathering front in some drier regions. Elsewhere the warmth, humidity and biological productivity have combined to produce younger saprolites with well-defined profiles.

Paradoxically, the most rapid weathering probably takes place in tectonic regions, where a combination of high rainfall, the occurrence of marine sediments (limestones, greywackes) and epithermal igneous rocks, plus stress fracturing of nearly all formations, promote weathering penetration and contribute to high erosion rates (Stallard 1995). However, the steep slopes erode rapidly and become weathering limited, deep weathering is therefore rare. In the humid tropics, steep terrain is subject to frequent landsliding, and the regolith becomes unstable when depths of 5–6 m are reached. As relief and slope are reduced, weathering profiles deepen and there is a need to research the thresholds governing this balance. Observations suggest that, where slopes are reduced below *c.*20°, weathering rates under forest can keep pace with the rate of regolith loss by erosion.

The phenomenon of deep weathering is, therefore, an expression of the formation and survival of materials in equilibrium with near-surface Earth environments. It involves the decay of minerals contained in rocks formed under pressure and in the absence of atmospheric gases, organic acids and micro-organisms, all of which are agents of chemical change. It is also an expression of the fluctuating rates of denudation in time and space. The great stores of saprolite that occur on the continents are, in some areas at least, relicts of the remote past, but weathering processes are continuous and deepening of the weathering mantle occurs where weathering is favoured and rates of denudation low.

References

Bird, M.I., and Chivas, A.R. (1988) Oxygen isotope dating of the Australian regolith, *Nature* 331, 513–516.

Eggleton, R.A. (ed.) (2001) *The Regolith Glossary*, Cooperative Research Centre for Landscape Evolution and Mineral Exploration (CRCLEME), CSIRO, Australia.

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

Fookes, P. (ed.) (1997) *Tropical Residual Soils*, Geological Society Professional Handbook, Bristol: The Geological Society.

Goldberry, R. (1979) Sedimentology of the Lower Jurassic flint clay bearing Mishor Formation, Makhtesh Ramon, Israel, *Sedimentology* 19, 229–251.

Hall, A.M. (1985) Cenozoic weathering covers in Buchan, Scotland and their significance, *Nature* 315, 392–395.

Lidmar Bergström, K. (1989) Exhumed Cretaceous landforms in south Sweden, *Zeitschrift für Geomorphologie NF Supplementband* 72, 21–40.

Löffler, E. (1977) *Geomorphology of Papua New Guinea*, Canberra CSIRO/Australian National University Press.

Mabbutt, J.A. (1965) The weathered land surface of central Australia, *Zeitschrift für Geomorphologie NF* 9, 82–114.

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

Millot, G. (1970) *Geology of Clays* (trans. W.R. Farrand and H. Paquet), London: Chapman and Hall.

Ollier, C.D. (1959) A two cycle theory of tropical pedology, *Journal of Soil Science* 10, 137–148.

——(1983) Weathering or hydrothermal alteration, *Catena* 10, 57–59.

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.

Sheppard, S.M.F. (1977) The Cornubian batholith, SW England: D/H and $^{18}O/^{16}$ studies of kaolinite and other alteration minerals, *Journal of the Geological Society of London* 133, 573–591.

Smith, B.J. and McAlister, J.J. (1995) Mineralogy, chemistry and palaeoenvironmental significance of an early Tertiary Terra Rossa from Northern Ireland: a preliminary review, *Geomorphology* 12, 63–73.

Stallard, R.F. (1995) Tectonic, environmental and human aspects of weathering and erosion, *Annual Review of Earth and Planetary Sciences* 23, 11–39.

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

Taylor, G.R., Eggleton, R.A., Holzhauer, C.C., Maconachie, L.A., Gordon, M., Brown, M.C. and McQueen, K.G. (1992) Cool climate lateritic and bauxitic weathering, *Journal of Geology* 100, 669–677.

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

MICHAEL F. THOMAS

DEFLATION

The process by which the wind removes fine material from the surface of a beach or desert. Large particles are left behind as a deflation lag and can cause ARMOURING and STONE PAVEMENT formation. Deflation from largely vegetation-free surfaces can create DUST STORMs and contribute to the formation of various features of wind erosion, including PANs and YARDANGs.

A.S. GOUDIE

DEGLACIATION

Deglaciation means the time period of uncovering of land or water by ice due to glacier retreat, normally forced by a climate change, in contrast to the term glaciation meaning the period of covering of land by ice. Deglaciation is related both to the major retreat of continental or regional-scaled glacier ice masses (large-scale deglaciation), especially during the glaciation phases in the Pleistocene, and for glacier shrinkage during the Holocene neoglaciation, e.g. since the Little Ice Age (small-scale deglaciation).

Deglaciation is triggered through climate changes (long-term) or climate variations (short-term) which impact the GLACIER mass balance due to changes in snow precipitation (accumulation) during winter and energy balance (e.g. temperature, radiation, latent heat release) during summer (ice melting). Increased summer energy input will increase ablation and cause an immediate response as retreat of the glacier front. The dynamic response of the glacier due to positive or negative mass balance of the glacier causes changes in the ice flux (mass transport) from the accumulation area down to the lower ablation area and will result in an advancing or retreating glacier front. The glacier front position will react on this forcing after a certain time period, known as the reaction time and the response time. The reaction time is given as the time lag between when the changes in mass balance occur and the first visible dynamic response of the front, and the longer response time defines the period until the glacier has stabilized to the new mass balance. These timescales are related to the dynamics of the glacier, and the glacier geometry, and can vary from a few years on a small valley glacier to several hundred or even thousands of years on large outlets from an inland ice sheet. The geometry and hypsometry (area-altitude distribution) of the glacier is important for the response. For example, if the Equilibrium Line Altitude (ELA) is raised by 100 m due to a warmer climate the increased area affected by more melting will be larger on a flat, wide glacier, and smaller on a steep, narrow glacier. Higher summer energy input, giving an immediate glacier front retreat and gradually lower ice transport over time, almost always causes deglaciation. During the last decades several energy balance models coupled to glacier-dynamical models have been developed, allowing the spatial and temporal simulation of glacier retreat due to different types and magnitudes of climatic forcing (e.g. Oerlemans 2001) (Figure 36).

The rate of deglaciation thus depends on climatic and topographic factors. Glaciers ending on land will usually get thinner and flatter during deglaciation. Calving glaciers will keep their steep, calving front, but get thinner and retreat much faster than glaciers ending on land. If the glacier is grounding in water, the reduction of mass flux may trigger buoyancy forces to lift up parts of the glacier front, which in turn leads to a massive up-calving of the glacier front. Such a rapid ice retreat occurred in deep fjord areas in western Scandinavia during the deglaciation of the Weichselian ice sheet (e.g. Sollid and Reite 1983).

The high temperature variability during the Pleistocene has caused numerous deglaciation phases in the northern hemisphere (Figure 37). According to present knowledge, there have been more than forty phases of glaciation and deglaciation during the Pleistocene. The last deglaciation was forced by a rapid increase of temperature. The mean Holocene temperature is about 10–13 °C

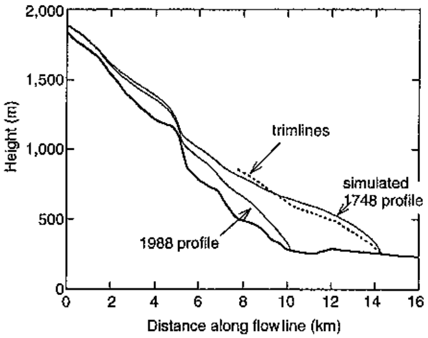

Figure 36 Glacier front positions at the outlet glacier Nigardsbreen, southern Norway (from Oerlemans 2001). The positions are obtained by applying a combined mass balance and dynamic glacier model

warmer than the mean temperature during full glacial conditions (Figures 37, 38). From the Greenland ice cores several rapid changes, called Dansgaard–Oeschger events, have been observed during the last glaciation (Dansgaard 1993). They all show an extremely rapid temperature increase of about 10° over only a hundred years followed by a slow cooling over several hundred years. At about 10,000 BP the temperature increased quickly again and stabilized at the Holocene temperature level, causing a rapid deglaciation. The forcing mechanisms for these large temperature changes over short time periods are discussed but not yet known. During Holocene the warmest period was in early to mid-Holocene and in many mountain regions the glaciers were probably melted away in the period 8,000–6,000 BP. The climate became colder from about 3,000 BP, starting the increase of glaciers in high mountain areas of the world (NEOGLACIATION), with a culmination in the period between the thirteenth century and about 1750 in Europe (e.g. Nesje *et al.* 2000), or about a hundred years later in some regions of the world. This period is known as the Little Ice Age. Since then deglaciation has prevailed until recent times in most glacial environments of the world. In some mountain regions with valley and cirque glaciers the mass loss has been massive, as for example in the Alps where the glacier retreat has resulted in a nearly 50 per cent reduction of the ice volume since the mid-1800s (e.g. IAHS(ICSI)/UNEP/UNESCO/ WMO 2001). Since the 1990s an accelerated deglaciation is observed in

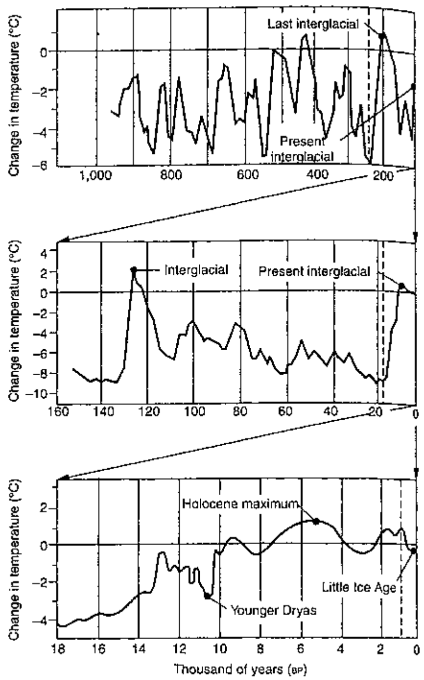

Figure 37 Temperature variation during the Pleistocene and Holocene, indicating glaciation and deglaciation phases (changed from Siegert 2001)

many alpine and in some arctic areas (Arendt *et al.* 2002; Meier and Dyurgerov 2002) and has been attributed to global warming.

Large-scale deglaciation led to a reduced weight on the landmasses, forcing a land heave. Simultaneously, the melting of glacier ice forced sea level to rise. Thus large-scale deglaciation is always related to land emergence and sea-level rise (see ISOSTASY; EUSTASY). During the maximum Weichselian ice extension global sea level was approximately 120 m lower than today. During the last deglaciation of the Weichselian ice sheet, the deglaciated area showed a net land heave, e.g. in Scandinavia. The central Bothnian area has emerged by more than 800 m since the deglaciation. Land heave produces continuously new coastlines with corresponding landforms, such as BEACH RIDGES and coastal ABRASION platforms, indicating the marine limits during a period of time. The spatial relationship between land heave

rates and marine limits has been used for relative dating (see DATING METHODS) of ice-recessional landforms. Furthermore, land heave results in subaerial exposure of the former sea bottom, covered by mainly marine clay-rich sediments. Areas covered by marine clays are abundant in areas of deglaciation, such as in Scandinavia and Canada. Having a high nutrient content and being easily arable these areas were of interest for early settlement and agriculture. However, these sediments are highly unconsolidated, and since deglaciation subject to severe GULLYing and prone to landsliding (see LANDSLIDE; QUICKCLAY).

The period of deglaciation is also the period of sediment accumulation by glacier ice and meltwater. Deglaciation is not a continuous process. Especially during the early phase of deglaciation the recession of ice margins frequently halted ('stagnation') or smaller re-advances occurred due to short-term climatic deterioration. The geological time periods of the Younger Dryas and the Preboreal are examples of such short-term climatic variations, which morphologically can be followed by glacial landforms almost throughout southern Finland, central Sweden and coastal Norway (Figure 38).

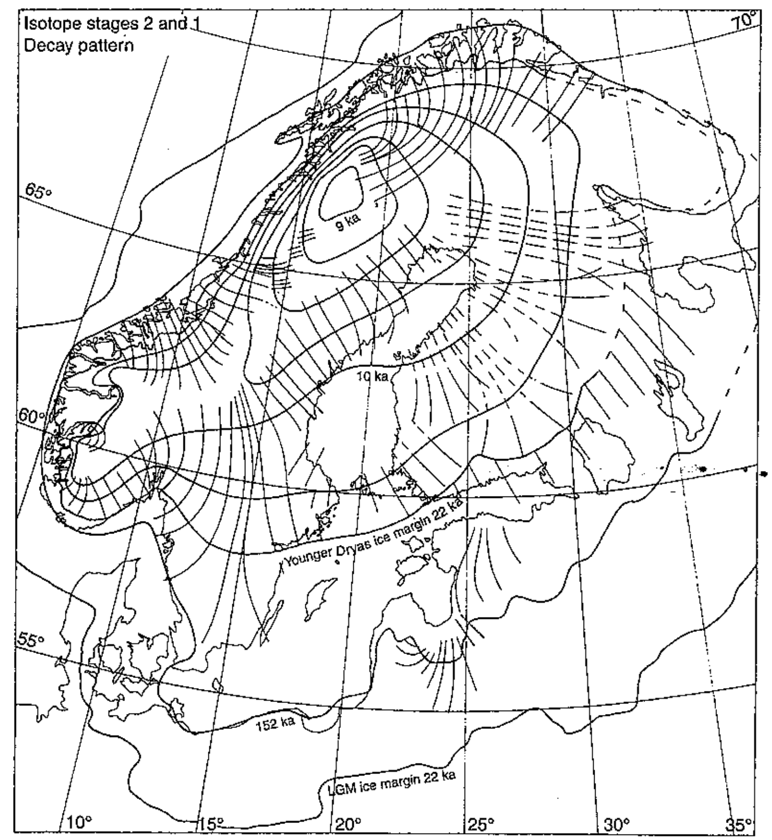

Figure 38 Glacial decay pattern of the Weichselian glaciation in Fennoscandia since the last glacial maximum (LGM, 22 ka) and the onset of the Holocene (*c.*10 ka) (adapted from Kleman *et al.* 1997)

The type of glacial land system (see Plate 33) during deglaciation phases is dependant upon whether the glacier front ends in water (marine/lacustrine) or on land (terrestrial), how fast the land is uncovered (deglaciation rate) and what glacier temperature regime prevailed during deglaciation (glacier thermal regime) (see also Benn and Evans 1998). A fast retreat of glacier tongues or lobes results in an uncovering of subglacial landforms such as fluted surface or drumlinoide forms. These subglacial landforms normally show the very last glacier movement direction. If the deglaciation happens in a permafrost environment, parts of the glacier marginal areas may be cold based. In such environments, glaciers may preserve sediments and landforms derived from earlier glaciation and deglaciation periods. Slow retreats and/or temporary stagnation of the glacier front produce landforms of accumulated glacigenic material, terminal moraines. During stagnation phases the glaciers advance some metres during winter and retreat during summer due to variations of ablation. The winter advances produce small annual moraines. Produced under water they are called DeGeer-moraines in marine environments and cross-valley moraines in mountainous lacustrine environments. Such moraines often build a sequence of landforms, used in deglaciation reconstruction. Glacier margins in permafrost environments are cold based. In this setting, a net-freezing condition along the glacier base prevails (e.g. Boulton 1972), forcing glacigenic material to accumulate in the glacier front area. Ice motion and surface melting leads then to an accumulation of glacigenic sediments on the glacier surface. If the thickness of this layer is larger than the ACTIVE LAYER thickness of the environment (see PERMAFROST), the ice below the layer is preserved, preventing further ablation. These ice-cored moraines can persist over long periods during deglaciation. If the active layer thickness increases or is eroded by e.g. fluvial action, the ice core melts and produces a hummocky moraine terrain proximal to the glacier front area and a clear distinctive border distal from the glacier. Deposited on steep terrain, the ice-cored moraines may continue creeping, producing ROCK GLACIER-like landforms. If the glacier front ends in water, most sediment transported in glacial meltwater is deposited in the vicinity of the glacier front, building up ice-contact deltas. When built up to the water surface, they give an indication of sea level during the deglaciation phase. Especially during the last stages of deglaciation, parts of the ice streams may become decoupled from the active part of the glaciers. This results in the loss of glacier flux and thus dead-ice wastage occurs. In this situation ice can be buried by e.g. glaciofluvial sediment. Melting of these ice bodies results in typical dead-ice landforms, consisting of hummocky irregular terrain and KETTLES AND KETTLE HOLES. On cold ice caps, meltwater is often routed along the glacier margins, forming channels that mark the ice surface during phases of deglaciation. Swarms of subsequently lower channels along mountain slopes indicate the lowering of the ice surface and the glacier surface slope during different phases of deglaciation.

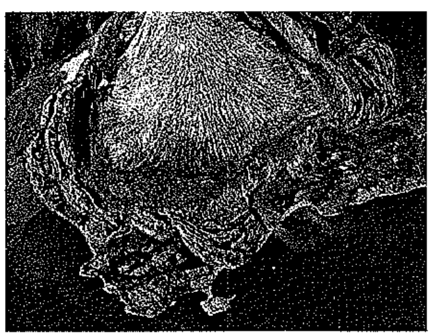

Plate 33 Oblique air photo of recently deglaciated terrain, Erikbreen, northern Spitsbergen (from Etzelmüller *et al.* 1996)

On continental ice sheets the ice divide did not necessarily correspond with the position of the topographic water divide of the underlying relief. During deglaciation topographic water divides often became ice free before all ice disappeared. This ice could occasionally block the drainage and thus formed ice-dammed lakes that drained over local or regional water divides. Like terminal moraines, lacustrine sediments and landforms such as shorelines bear witness to periods of deglaciation. Sudden outbursts of glacial lakes are called by the Icelandic word *jokulhaup*. In many high mountain areas, e.g. in central Asia, deglaciation of valley glaciers leads to damming of lakes between the glacier front and terminal moraines, which often are ice-cored. These lakes are unstable, and outbursts, often called GLOFs – glacier lake outburst floods – are a potential hazard for lower lying valleys and human settlements and infrastructures. The same risk applies to the situation where valley glaciers block the drainage from minor side valleys.

Research on deglaciation is concentrated on (1) determining the start of deglaciation, (2) the deglaciation rate, and (3) the change of the spatial distribution of the ice body during different phases of deglaciation. Traditionally, scientists concentrated on determining the start and deglaciation rate, by dating of landforms associated to terrestrial or marine/lacustrine glacier margins (see DATING METHODS) and analysing sediment succession building up these landforms. Coring of marine sediments on continental margins and deep ocean basins revealed continuous information on glaciation and deglaciation phases during the Pleistocene (e.g. Elverhøi *et al.* 1995). The Holocene NEOGLACIATION chronology is depicted by core analysis of local sediment sinks such as lakes fed by meltwater from glaciers (e.g. Karlén 1976). The spatial distribution of ice bodies during deglaciation phases is often determined through GEOMORPHOLOGICAL MAPPING of glacial landforms combined with dating methods. The past vertical extent of an ice sheet in its accumulation area is difficult to obtain because of the lack of marked landforms due to low glacier velocities and often cold ice in culmination zones of ice sheets. Recently, exposure dating using cosmogenic isotopes has proved to be a helpful tool in this respect.

Deglaciation has become an important problem for human settlement and sustainable development in many high mountain environments. Especially in many semi-arid areas, such as the eastern slopes of the Andes Mountains and in central Asian mountain ranges, glaciers act as freshwater reservoirs, since meltwater from the glaciers is important for irrigation and water supply. Deglaciation leads to a periodically enhanced runoff. However, due to shrinkage of the water reserve (the glaciers), water availability will be reduced over the long term.

References

Arendt, A.A., Echelmeyer, K., Harrison, W.D, Lingle, C.S. and Valentine, V.B. (2002) Rapid wastage of Alaska glaciers and their contribution to rising sea level, *Science* 297, 382–386.

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

Boulton, G.S. (1972) The role of thermal regime in glacial sedimentation, *Institute of British Geographers. Special Publication* 4, 1–19.

Dansgaard, W.E.A. (1993) Evidence for general instability of past climate from a 250-kyr ice-core record, *Nature* 364, 218–220.

Elverhøi, A., Anderson, E.S., Trond, D., Hebbeln, D., Spielhagen, R.F., Srendsen, J.I. *et al.* (1995) The growth and decay of late Weichselian ice sheet in western Svalbard and adjacent areas based on provenance studies of marine sediments, *Quaternary Research* 44, 303–316.

Etzelmüller, B., Hagen, J.O., Vatne, G., Ødegård, R.S. and Sollid, J.L. (1996) Glacier debris accumulation and sediment deformation influenced by permafrost, examples from Svalbard, *Annals of Glaciology* 22, 53–62.

IAHS(ICSI)/UNEP/UNESCO/WMO (2001) *Glacier Mass Balance, Bulletin No. 6*, Haeberli, W., Frauenfelder, R. and Hoelzle, M. (eds) World Glacier Monitoring Service, University of Zurich and ETH Zurich.

Karlén, W. (1976) Lacustrine sediments and tree-limit variations as indicators of Holocene climatic fluctuations in Lappland, northern Sweden, *Geografiska Annaler* 58A, 1–34.

Kleman, J., Hättestrand, C., Borgström, I. and Stroeven, A.P. (1997) Fennoscandian palaeoglaciology reconstructed using a glacial geological inversion model, *Journal of Glaciology* 43(144), 283–299.

Meier, M. and Dyurgerov, M. (2002) How Alaska affects the world, *Science* 297, 350–351.

Nesje, A., Dahl, S.O., Andersson, C. and Matthews, J.A. (2000) The lacustrine sedimentary sequence in Sygneskardvatnet, western Norway: a continuous, high-resolution record of the Jostedalsbreen ice cap during the Holocene, *Quaternary Science Reviews* 19, 1,047–1,065.

Oerlemans, J. (2001) *Glaciers and Climate Change*, Lisse: A.A. Balkema.

Siegert, M.J. (2001) *Ice Sheets and Late Quaternary Environmental Change*, Chichester: Wiley.

Sollid, J.L. and Reite, A. (1983) Central Norway, glaciation and deglaciation, in J. Ehlers (ed.) *Glacial Deposits in North-west Europe*, 41–59, Rotterdam: Balkema.

SEE ALSO: dating methods; eustasy; glacier; ice dam, glacier dam; isostasy; moraine; neoglaciation

BERND ETZELMÜLLER AND JON OVE HAGEN

DELL

Small headwater valleys which are characteristically sediment-choked and swampy. Dells frequently occur at the head of deep gorges on plateaux surfaces and may be analogous to DAMBOS. Notable dells have developed on sandstone on the Woronora Plateau of New South Wales, Australia (Young 1986).

They are also known from Eocene beds in the New Forest of southern England, where they may have a periglacial origin and form tributaries to small dry valleys (Tuckfield 1986).

References

Tuckfield, C.G. (1986) A study of dells in the New Forest, Hampshire, England, *Earth Surface Processes and Landforms* 11, 23–40.
Young, A.R.M. (1986) The geomorphic development of dells (upland swamps) on the Woronora Plateau, N.S.W., Australia, *Zeitschrift für Geomorphologie NF* 30, 317–327.

A.S. GOUDIE

DEMOISELLE

A French term used to describe a needle-shaped Earth pillar composed of eroded rock, and capped by a large boulder. The overlying block is typically more resistant than the underlying material, and so tends to protect it from erosion (predominantly by water), as well as helping to maintain its vertical integrity. Demoiselles are commonly found in Alpine areas of highly weathered volcanic breccia or of glacial till. The term is derived from French for 'young lady', and is commonly employed throughout the French Alps. The term *cheminée de fees* (fairies' chimney) is also used frequently in place of demoiselle, while the American synonym of demoiselle is HOODOO.

STEVE WARD

DENDROCHRONOLOGY

Dendrochronology is the study of annual tree rings, with studies based on the measurement of variations in ring widths caused by variations in climate and environment at the time of ring formation. Ring counts and ring-width measurements provide precise calendar dating for ring formation years and a basis for numerous research applications.

The main disciplines included in these applications are: dendroclimatology, dendroarchaeology, art historical research, history, dendroecology and the related disciplines of DENDROGEOMORPHOLOGY, dendroglaciology, dendrohydrology, dendroniveology (snow and ice research), dendropyrology (fire events) and dendroseismology. Tree-ring chronologies have also been used to calibrate the radiocarbon timescale.

Douglass explored the potential of tree rings for climate analysis in 1919, but it was not until 1953 with the discovery of the 4,000-year old Bristlecone pines that the technique drew widespread attention. By the end of the century, chronologies covering nearly 10,000 years had been constructed from matching ring-width patterns of living and dead trees, and dendrochronology was being applied worldwide in research on global change, although as yet work in the tropics has been limited.

Procedures entail either cutting discs from a stem or, more usually and less destructively, the collection of wood cores with a 5 mm increment borer. The core is usually glued to a wood support, with its grain perpendicular to the support, and polished to reveal the ring structures ready for ring-width measurement.

Coring causes mechanical injury. Thus, it is important to obtain permission before taking samples and never to core anything that could be valuable as timber. However, trees compartmentalize wounds and often produce anti-fungicidal substances that generally limit damage; injuries stimulate local growth and holes are callused over in a few years. Studies have indicated that the core hole should be left open and untreated since this could introduce foreign organisms and impede the healing process.

An annual tree ring usually has two growth phases. At the beginning of the growing season conifers produce large, pale, thin-walled earlywood cells; towards the end of the season increasingly small diameter, dark, thick-walled, latewood cells develop. Hardwood trees have a variety of ring forms. Healthy, unstressed trees will produce concentric rings with approximately equal ring widths while stressed trees will form eccentric ring patterns and show narrow, variable ring growth. Trees on slopes are frequently bent, with deciduous species having their central pith displaced towards the downslope side of the stem and conifers to the upslope side (a point to be remembered when coring bent trees).

Problems for the technique, apart from those caused by growth eccentricities, are introduced by non-uniform cell growth due to adverse conditions, with normal cell formation either halted or present over only part of a stem resulting in missing rings. Alternatively, false rings can be produced by late frosts, droughts or other growth-inhibiting events resulting in darkened cells followed by resumption of normal growth before true dark latewood growth marks the end of the growing season. These complications can be mitigated by crossdating.

Crossdating is achieved by matching sample ring-widths using visual and statistical tests. At least two cores are usually collected from each tree, so that they can be crossdated to check for missing or false rings and, where there are none the radii are averaged to show mean annual ring growth. Graph plots of the means of individual trees are then compared and crossdated and a site masterplot created.

Sample depth (number of trees sampled) will change through time affecting the quality of a chronology. Consequently, chronologies may be truncated where there are less than three trees to support a mean curve and, for valid climatic results, curves should contain an absolute minimum of ten trees per site, but thirty or more are desirable. Where sample depth is important, this information should be included on ring-width graphs.

The technique was revolutionized by the advent of computer processing enabling the digitizing of ring-width measurements; rapid plotting and comparison of graphs; rapid application of multi-variant statistics, and radio-densitometric determination of wood density. This latter approach is based on X-ray analysis of changes in cell densities; it is used particularly in climate studies to highlight sensitive reactions of cell densities to temperature variations. It is also used for analysis of tropical species' growth, since these species, rather than always forming annual rings, may produce growth zones reflecting aperiodic precipitation or drought.

Prior to computerization, measurements were made by hand and one approach to crossdating was the use of 'skeleton plotting' based on ring counting, visual assessment of relative per cent ring widths, and identification of event years. Skeleton plots provide dating for, and a visual summary of, the effects of environmental events without the need to make precise ring-width measurements. It is a useful procedure showing major events and growth trends where rapid assessment of limited sample numbers is required.

Apart from the effect of sudden events on tree growth, slow growth changes may occur due to gradual variations in climate or natural reduction in ring width as a tree ages. This latter effect is routinely removed by standardizing (detrending) ring widths using various techniques. Standardization, apart from eliminating age trends, emphasizes event years while removing climatic trends shown by moving averages of the mean ring-width data.

Further reading

Cook, E.R. and Kairiukstis, L.A. (eds) (1990) *Methods of Dendrochronology: Applications in the Environmental Sciences*, Dordrecht: Kluwer Academic Publishers.
Esper, J. and Gärtner, H. (2001) Interpretation of tree-ring chronologies, *Erdkunde* 55, 277–288.
Schweingruber, F.H. (1989) *Tree Rings: Basics and Applications of Dendrochronology*, Dordrecht: Kluwer Academic Publishers.
Schweingruber, F.H., Eckstein, D., Serre-Bachet, F. and Bräker, O.U. (1990) Identification, presentation and interpretation of event years and pointer years in dendrochronology, *Dendrochronologia* 8, 9–38.

VANESSA WINCHESTER

DENDROGEOMORPHOLOGY

Dendrogeomorphology is based on analysis of the annual growth rings of trees or woody plants and their growth forms. It is used to investigate spatial and temporal aspects of Earth surface processes operating during the Holocene at annual to centennial timescales.

The technique is closely associated with dendroclimatology and employs largely the same methods as DENDROCHRONOLOGY. Applications include dating and establishing rates of change and frequencies of storms, FLOODS, LAKE outbursts, river channel changes, frost events and ICE surges, GLACIER movements, snow avalanches (see AVALANCHE, SNOW) MASS MOVEMENTS and FIRES, and show the relationships of events to climate. In addition, tree rings can provide records of events unrelated to climate: volcanic eruptions; earthquakes and TSUNAMIS; environmental management; COMPACTION OF SOIL; water table variations, changes in pollution and saltwater ingression.

Methods, other than those used in dendrochronology, include studies of the age, anatomy, morphology, and structures of tree roots, stems and crowns. Root ring patterns can be used to date sediment aggradation or degradation. Trees respond to increases in soil depth by producing adventitious roots; soil movements cause root structures to bend while degradation leaves roots exposed. Ring counts supply dates for root structures, bends and stem age at ground level while age and distances between features show the scale of events. Eccentric ring patterns develop where roots are part-exposed or when denudation brings them close to the surface. Changes in patterns supported by changes in cell structures can be dated. Before sampling buried or exposed roots, records should be made of all relevant features: positions and orientations of main and adventitious root systems; distances to the ground surface; vegetation cover and soil type.

Stems deformed by site changes produce eccentric ring patterns. Stem discs or cores taken both in the direction of stress and at right angles, show when eccentricity begins and the orientation of patterns provides information on the direction of changes. Injuries to stems or roots produce scarring with local growth being stimulated. A core from an undamaged area near the wound (but avoiding re-growth tissue) will show the number of years elapsed since the event.

Crown development provides information on competition, wind and storm events, snow cover and tree health.

The main problems for dendrogeomorphology where surface age is the focus of interest are to establish the total age of a tree and the length of time taken for a tree to colonize a freshly exposed surface. Core ring counts only show the age of a tree above the coring point; thus to find total age an estimate of the number of years growth below this is required. One method is to cut stem discs near ground level of a number of small trees growing in a range of local microenvironments, correlate the height of the trees with their age and calculate the mean growth to height ratio for the location. Verification of ecesis (colonization) times requires an alternative dating source.

Further reading

Gärtner, H., Schweingruber, F.H. and Dikau, R. (2001) Determination of erosion rates by analysing structural changes in the growth pattern of exposed roots, *Dendrochronologia* 19, 81–91.

Schweingruber, F.H. (1996) *Tree Rings and Environment Dendroecology*, Berne: Paul Haupt.

Strunk, H. (1997) Dating of geomorphological processes using dendrogeomorphological methods, *Catena* 31, 137–151.

Winchester, V. and Harrison S. (2000) Dendrochronology and lichenometry: an investigation into colonization, growth rates and dating on the east side of the North Patagonian Icefield, Chile, *Geomorphology* 34, 181–194.

VANESSA WINCHESTER

DENUDATION

On Earth, two forces counterbalance: uplift (i.e. creation of relief) and denudation. Denudation includes all processes that remove the relief at the surface of the Earth. Denudation acts chemically or physically. Chemical denudation, also termed chemical weathering or chemical erosion, is the slow complete or partial dissolution of rock minerals. Physical denudation or mechanical weathering processes correspond to the removal of solids from the land surface. Quantification is generally expressed by the mean of chemical or physical fluxes (or rates) of denudation, expressed most often in $t\,km^{-2}\,a^{-1}$.

The adaptation of rock minerals to the conditions at the surface of the Earth (Ahnert 1996) releases the most soluble elements and leads generally to the formation of residual minerals, usually clays and hydrous iron or aluminium oxides, that accumulate at the interface between the atmosphere and the lithosphere (the REGOLITH). Examples of chemical denudation reactions are the weathering of calcite, which does not leave any residue

$$CaCO_3 + H_2O + H^+ = HCO_3^- + Ca^{2+}$$

and the weathering of albite, that produces a secondary phase, for example kaolinite:

$$2NaAlSi_3O_8 + 9H_2O + 2H^+ = Al_2Si_2O_5(OH)_4 + 2Na^2 + 4H_4SiO_4$$

These equations show that protons are necessary to attack rock minerals. At the surface of the Earth, these protons are mostly derived from the dissolution of the atmospheric or soil CO_2 in water

$$CO_2 + H_2O = H_2CO_3 = H^+ + HCO_3^-$$

Weathering reactions therefore pump CO_2 from the atmosphere and convert it into bicarbonate ions. Ultimately, these ions, combined with the Ca and Mg ions liberated by rock weathering, will lead to the precipitation of calcite in the ocean by living organisms, allowing the sequestration of atmospheric-derived carbon. Other possible origins for protons include production by organic molecules derived from the degradation of organic matter in soils (humic and fulvic acids) and the oxidation of sulphide minerals producing sulphuric acid.

Like minerals, rocks do not weather at the same rate. A good way of estimating the rate of chemical weathering is by analysing rivers draining a single type of rock, provided that the river dissolved load is corrected from the inputs that do not derive from rock weathering (atmosphere, pollution, biomass). Another approach is based on soil mass budgets. Rates of chemical denudation

are extremely variable, ranging from less than $t\,km^{-2}\,a^{-1}$ in high latitude granitic catchments or in the low-lying regions of central Africa, to more than $100\,t\,km^{-2}\,a^{-1}$ for rivers draining basalts at Réunion or Java Island (Louvat and Allègre 1997). Basaltic lithologies thus weather 10 to 100 times faster than granites. At a global scale, it has been shown by Dessert *et al.* (2002), that, even if the outcrops of basalts represent 5 per cent of the emerged surface of the Earth, the flux of CO_2 uptake by basalt weathering is as high as 35 per cent of the total consumption flux by rock weathering. Basalt weathering therefore appears as a major mechanism of atmospheric CO_2 regulation. Carbonate rocks also have high denudation rates, ranging from 10 to $200\,t\,km^{-2}\,a^{-1}$. Saline rocks have the highest chemical denudation rates because they are highly soluble in water. From large river systems, chemical denudation rates (Figure 39) ranging from a few $t\,km^{-2}\,a^{-1}$ for the Zaire, Nile and Siberian rivers to about $50\,t\,km^{-2}\,a^{-1}$ for rivers such as the Mekong, Mackenzie or Brahmaputra have been determined (Summerfield and Hulton 1994). These rates are strongly correlated to the abundance of carbonates within the drainage basin, simply because carbonates and evaporites weather at a faster rate than silicates.

Although lithology is the first controlling factor on chemical weathering rates, other parameters exert a control on chemical denudation. At both small and large scales, chemical weathering rates increase strongly with runoff and temperature. This is especially true for basalt weathering rates which respond at a global scale to an Arrhenius-type law (Dessert *et al.* 2002). At a continental scale, however, several authors have pointed out that the highest chemical denudation rates of silicate rocks are not found in the regions of highest rainfall and temperatures (Edmond *et al.* 1994). The Zaire river has the same chemical denudation rates as the Yenisey (Gaillardet *et al.* 1999). The low chemical denudation rates found in the flat and humid tropical areas contrast with the highly weathered nature of soil material (laterites) that characterize these regions. At a global scale, intensity and flux of chemical denudation of silicates are inversely correlated. This paradox will be explained later.

Physical denudation rates can be estimated by different means: by using rivers, sediment accumulation in reservoirs or sedimentary basins and

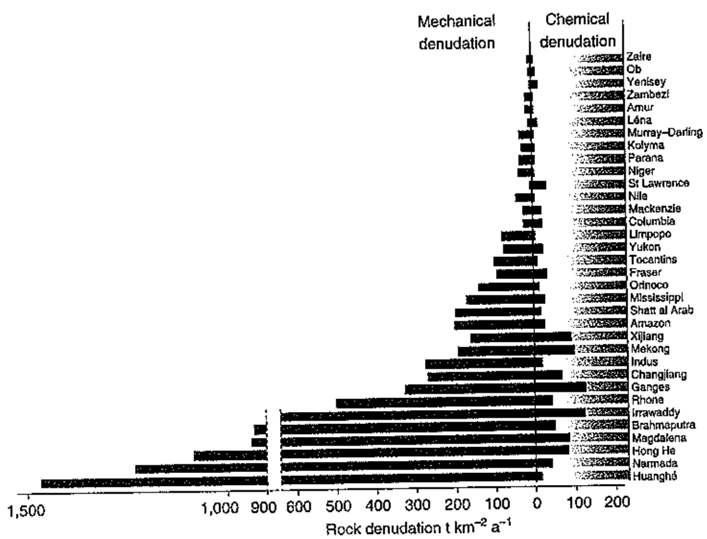

Figure 39 Physical vs chemical denudation rates for the world's largest river basins. Solute load is corrected from atmospheric inputs. Based on Summerfield and Hulton (1994) and Gaillardet *et al.* (1999)

cosmogenic isotopes. Rivers are probably the easiest way of calculating mechanical denudation rates, but this approach suffers from various limitations. Unlike dissolved load, which often fluctuates by a factor of two or so between high water and low water stages, the quantity of solids transported by rivers can vary drastically. A single daily flood event can transport as much sediment as several years of regular sediment flux. In addition, the amount of material transported as bottom sand is generally unknown. Rivers, especially large rivers, store sediments in their floodplains or alluvial fans. For example, it has been shown that two-thirds of the sediments removed from the Andes by the headwaters of the Rio Madeira in South America are stored in the foreland floodplains and never reach either the Amazon, or the sea (Guyot 1993). Finally, the influence of anthropogenic activities has usually resulted in a strong increase of river sediment yield. The best example is the Huanghe river, that transports, today, up to $20\,\mathrm{g\,l^{-1}}$ of sediments to the ocean, while the long-term pre-anthropogenic estimate of Holocene sediment volumes deposited by the river indicates an average suspended load concentration one order of magnitude lower (Milliman and Syvitski 1992).

Mechanical denudation rates (Figure 39) have been computed for the largest river systems by a number of authors (Pinet and Souriau 1988; Milliman and Syvitski 1992; Summerfield and Hulton 1994). They range from numbers lower than $10\,\mathrm{t\,km^{-2}\,a^{-1}}$ for rivers such as the Siberian (Ob, Yenisey, Kolyma) and tropical (like the Zaire), to numbers higher than $700\,\mathrm{t\,km^{-2}\,a^{-1}}$ for the largest rivers of Asia (Brahmaputra, Ganges). The world average value is estimated to be $200\,\mathrm{t\,km^{-2}\,a^{-1}}$, corresponding to $20.10^9\,\mathrm{t\,a^{-1}}$ (Milliman and Syvitski 1992).

The dominant factors that influence mechanical denudation are the erodibility of rocks and relief. High relief areas tend to have high mechanical denudation rates. This is mainly due to slope instabilities and to glacial abrasion. There does not seem to exist any clear relationship between climate (runoff or temperature) and physical erosion, at least on a global scale (Pinet and Souriau 1988; Summerfield and Hulton 1994). Regions of high precipitation and high seasonality of rainfall seem however to exhibit higher mechanical denudation rates (see Goudie 1995). An inverse correlation between suspended yields and basin area is reported by an extensive study of Milliman and Syvitski (1992), possibly showing the importance of sediment storage. The same authors showed that humans are also a major controlling parameter, as fluvial denudation rates during the Holocene are estimated to be less than half the present-day rates. However, for the largest rivers of Asia, there is a remarkably good agreement between present-day fluvial physical denudation rates and long-term rates based on sediment volume (Métivier and Gaudemer 1999) accumulated in the sea. Based on the denudation rates computed on large rivers, Gaillardet *et al.* (1999) have shown that chemical denudation rates of silicate rocks are positively correlated to physical denudation rates. Such a relation is confirmed by cosmogenic nuclides measurements (Riebe *et al.* 2001). This global coupling between chemical and physical fluxes of silicate denudation is explained as follows. With a low mechanical denudation regime, soil development is favoured, leading to a shielding effect of the soil and a negative feedback on the interaction between water and mineral surfaces. This is the transport-limited regime. Conversely, in regions of high mechanical denudation mineral surfaces are continuously exposed, and even if chemical weathering is not intense, the fluxes of released solutes are increased. This is the weathering-limited regime. Overall, the present-day Earth is under the weathering-limited regime.

Total denudation rates are calculated by adding the chemical and physical denudation rates. Using a mean crustal density of $2{,}700\,\mathrm{kg\,m^{-3}}$, these rates are usually translated in $\mathrm{mm\,a^{-1}}$ (Figure 40). For large rivers, landscape downwearing ranges from about $10\,\mathrm{mm\,a^{-1}}$ in regions such as the shield low-lying areas of the Congo craton, Niger basin, Brazilian shield and Australian shield to 100–200 mm per 1,000 years for the Ganges, Indus, Changjiang and Amazon rivers. The Brahmaputra river has the highest rate of total denudation, close to 700 mm per 1,000 years. The mean value for the Earth's surface is 61 mm per 1,000 years (areas of internal drainage excluded). Total denudation rates calculated by mass budgets of riverine products are in good agreement with cosmogenic isotope measurements. Cosmogenic isotopes are produced during the bombardment of cosmic rays and their abundance is a function of the production rate and total erosion rate. This technique has been applied by a number of authors and gives denudation rates, integrated over tens of millions of years in the case of ^{10}Be, which are in general agreement with other techniques. For example, the ^{10}Be derived denudation rates of the Loire, Meuse, Neckar and Regen basins are in the order of magnitude of those found by conventional techniques (Schaller *et al.* 2001).

As a global average, 10–15 per cent of the total denudation is chemical denudation, 75–80 per cent is physical denudation (Figure 39). The ratio of mechanical over chemical fluxes fluctuates widely from about 1 for the Siberian and tropical rivers to 10–20 for mountainous rivers. The present world is therefore dominated by physical denudation, but it may not have been the case for geological periods of low relief.

Atmospheric dust transport from the continent to the ocean is also responsible for the denudation of continents. Aeolian denudation is extremely difficult to quantify and estimates vary between 0.1 to $5.10^9\,\mathrm{t\,a^{-1}}$. In deserts, aeolian denudation may be the only process of relief denudation.

References

Ahnert, F. (1996) *Introduction to Geomorphology*, London: Arnold.

Dessert, C., Dupré, B., Gaillardet, J., Francois, L. and Allègre, C. (2003) Basalt weathering laws and global geochemical cycles, *Chemical Geology*. In press.

Edmond, J.M. *et al.* (1994) Fluvial geochemistry and denudation rate of the Guyana shield, *Geochimica et Cosmochimica. Acta* 59, 3,301–3,325.

Gaillardet, J., Dupré, B., Louvat, P. and Allègre, C. (1999) Global silicate weathering of silicates estimated from large river geochemistry, *Chemical Geology* 159, 3–30.

Goudie, A. (1995) *The Changing Earth. Rates of Geomorphological Processes*, Oxford: Blackwell.

Guyot, J.L. (1993) Hydrogéochimie des fleuves de l'Amazonie bolivienne, Ph.D. thesis, Université Bordeaux I.

Louvat, P. and Allègre, C.J. (1997) Present denudation rates at Réunion island determined by river geochemistry: basalt weathering and mass budget between chemical and mechanical erosions, *Geochimica et Cosmochimica. Acta* 61(17), 3,645–3,669.

Métivier, F. and Gaudemer, Y. (1999) Stability of output fluxes of large rivers in South and East Asia during the last two million years. Implications on floodplain processes, *Basin Research* 11, 293–304.

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

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

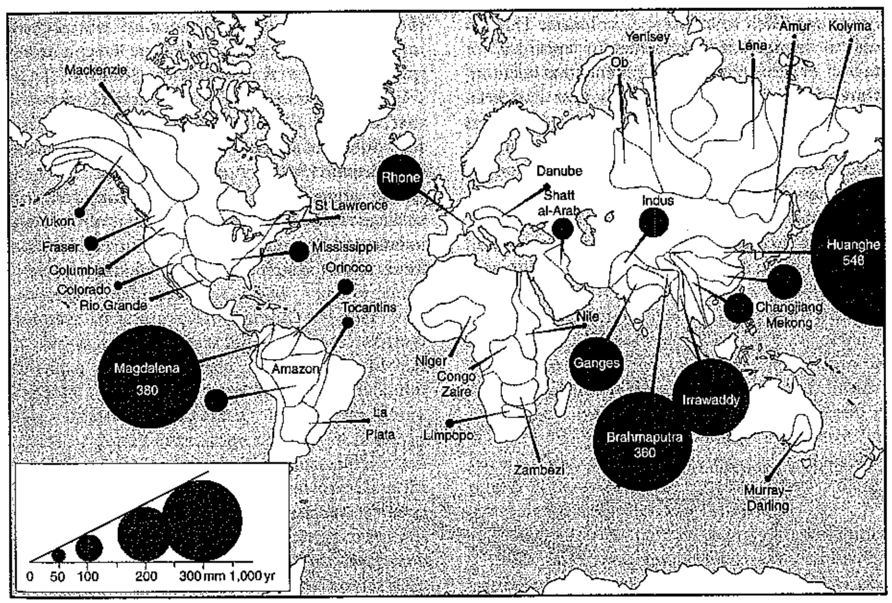

Figure 40 Total denudation rates for the largest drainage basins calculated with a mean density of $2.7\,\mathrm{g\,cm^{-3}}$. Based on Summerfield and Hulton (1994) and Gaillardet *et al.* (1999)

Riebe, C.S., Kirchner, J.W., Granger, D.E. and Finkel, R.C. (2001) Strong tectonic and weak climatic control of long-term chemical weathering rates, *Geology* 29, 511–514.
Schaller, M., von Blanckenbourg, F., Hovius, N. and Kubik, P.W. (2001) Large-scale erosion rates from in situ-produced cosmogenic nuclides in European rivers, *Earth and Planetary Science Letters* 188, 441–458.
Summerfield, M.A. and Hulton, N.J. (1994) Natural control of fluvial denudation rates in major world drainage basins, *Journal of Geophysical Research* 99, 13,871–13,883.

Further reading

Summerfield, M.A. (1991) *Global Geomorphology*, London: Longman; New York: Wiley.

SEE ALSO: cosmogenic dating

JÉRÔME GAILLARDET

DENUDATION CHRONOLOGY

The explanation of how topographic landscapes came to attain their present form has always been a prime objective of geomorphology. Prior to the 1960s, most workers adopted an essentially historical approach to landscape evolution. The aim was to identify the sequence of episodes, or stages, of erosional development that demonstrated how contemporary landscapes had been sculptured from hypothetical initial topographies that were usually considered fairly uniform and featureless. This sequential approach to topographic evolution, with its focus on DENUDATION, came to be known as denudation chronology. A significant proportion of such studies were based on the Davisian model of landscape evolution and can be termed classical denudation chronology, an historic element of a broad field of study currently known as long-term landscape development or evolutionary geomorphology.

Classical denudation chronology sought to identify evidence of past PLANATION SURFACEs and erosional levels in a landscape and to place them in a chronological sequence. Two key concepts were employed. First, that topographic 'flats', bevels and benches, together with accordant ridge and summit levels, represented the remnants of marine platforms, SUBAERIAL low-relief surfaces or terraces produced during past periods of relatively stable BASE LEVEL or 'stillstand'. Second, that there had been a progressive but episodic fall in base level with time, so that the highest features (in terms of elevation) were the oldest. The resulting 'geomorphological staircases' often rose via terraces and benches to the more fragmentary remains of 'summit surfaces' preserved on ridges and CUESTAS (e.g. the 'Schooley Peneplain' of W.M. Davis (Figure 41), the 'Mio-Pliocene Peneplain' of Wooldridge and Linton (Figure 42)), or to even higher surfaces whose former existence was postulated on the basis of the summits of residual hills or 'monadnocks' (Figure 42), or by the projection of the planes of outcropping unconformities (e.g. the 'Fall Zone Peneplane' rising above the Appalachians (Johnson 1931) (Figure 41), the 'Sub-Eocene Surface' of south-east England (Wooldridge and Linton 1955) or the 'Sub-Cretaceous Surface' rising over Wales (e.g. Brown 1960)).

The identification and delimitation of such surfaces was usually based on visual observation, augmented by field mapping, profiling and various kinds of cartographic analysis, including the use of superimposed and projected profiles (Figure 42), and the resulting sequences could be extended back into the geological past by use of the basic rules of stratigraphy, such as the laws of superposition and of original horizontality and the nature of unconformable relationships. Hence many denudation chronologies were extended through the Neogene into the Palaeogene and even into the Mesozoic (Figure 41). Although this was often claimed to be a strength, because such studies provided a bridge between Quaternary landscape development and stratigraphy, in reality the resultant 'histories' had much closer affinities with historical geology than with geomorphology, especially once contemporary process studies had developed to dominate geomorphology.

Relatively little emphasis was placed on the study of surficial deposits, largely because of the limited availability of analytical techniques and the complexity of such deposits. Often the studies that were undertaken focused on whether or not the identified erosional remnants were of subaerial or marine origin. By contrast, drainage patterns and drainage evolution figured prominently. Drainage-structure relationships, together with the existence of cols, wind-gaps and UNDERFIT STREAMs, were all used to recreate former drainage patterns associated with particular stages of landscape evolution (see Brown 1960). Often the aim of such studies was to recreate the original pattern of sub-parallel consequent rivers that developed on an uplifted and tilted marine plane or were superimposed from overlying cover strata across a fundamental unconformity. The identification and interpretation of discordant drainage therefore became a highly contentious issue and led to great debates about the extent and significance of former marine planation surfaces (see Jones 1981).

It is often assumed that classical denudation chronology evolved from W.M. Davis's exposition of the concept of the CYCLE OF EROSION, but this is incorrect. Long before the Davisian model was first outlined in 1889 others had begun to develop simple, embryonic denudation histories. For example, in Britain there was the work of Ramsay in 1846 and 1864, Jukes in 1862 and Topley in 1875, while in America McGee in 1888 developed an erosional chronology for that very same part of the Appalachians that was to be made classic by W.M. Davis's own detailing of the cycle of erosion concept (Davis 1889).

The Davisian model, as subsequently refined and elaborated (see Davis 1909), with its notions of peneplanation and cyclic change due to variations in base level, clearly fitted in so well with notions of the sequential development of landscapes, that it is no surprise that the two approaches merged. Many high-level marine surfaces were reinterpreted as PENEPLAINs and every attempt was made to place identified 'flats' into discrete groupings on the basis of elevation (Figure 43) in order to recreate cycles and partial cycles which could then be correlated with evidence from adjacent regions on the basis of elevation alone. Thus the cycle of erosion concept invigorated denudation chronology and, in turn, the concept was to survive as a basic element of classical denudation chronology long after it had been discarded by the remainder of geomorphology, following advances in process-based understanding of landform development.

During the first half of the twentieth century, denudation chronology became a major preoccupation of geomorphological studies: in America, under the influence of D.W. Johnson (1931), in Britain, where S.W. Wooldridge was the dominant figure (see Wooldridge and Linton 1955)

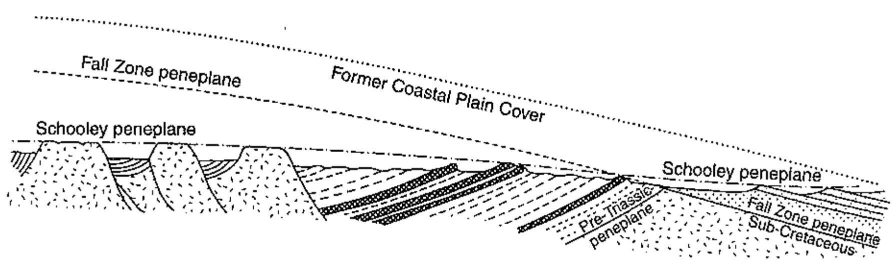

Figure 41 Scheme of landscape development for the eastern USA first advanced by D.W. Johnson in 1928

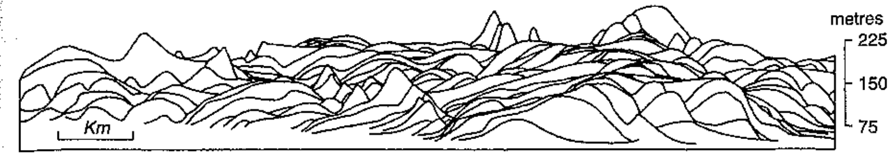

Figure 42 Projected profiles on the Chalk, western margins of Salisbury Plain, southern England, published by C.P. Green in 1974. The residual hills indicate the altitude of the former Sub-Eocene (now Sub-Palaeogene) surface and rise above the Mio-Pliocene Peneplain of Wooldridge and Linton, which Green subdivided into Summit Peneplain: Higher Surface and Summit Peneplain: Lower Surface, dated as Miocene and Pliocene respectively (see Jones 1981)

and in France, where the pioneering study of the Massif Central by H. Baulig (1928) established the blueprint for later studies in Europe. However, there were differences across the Atlantic regarding the emphasis placed on eustatic versus diastrophic mechanisms as the main cause of base level changes, with European studies following the lead of Baulig in favouring eustatic change. This is classically displayed in B.W. Sparks's (1949) interpretation of the morphology of the South Downs cuesta backslopes in southern England as consisting of eight marine levels between the two proven raised beaches below 40 m and the postulated Plio-Pleistocene marine plain at 170–200 m (Figure 43) – a geomorphological staircase encompassing the Pleistocene in no fewer than eleven treads, the majority of which lacked supporting sedimentological evidence.

Denudation chronologies were developed for other areas using modified models of landscape evolution, most dramatically in the case of South Africa where Lester King (1972) produced a classic sequential interpretation of the area between the Drakensberg Escarpment and the Natal Coast (Figure 44), including artistic representations of the landscapes developed at each stage. King's model of landscape evolution represents an amalgam of the views of Davis and Penck; episodic uplift resulting in both downwearing and backwearing, with the parallel retreat of slopes leading to the formation of PEDIMENTS which coalesced to form pediplains through the process of pediplanation. King himself went so far as to state that the morphological and sedimentological sequences identified in Natal provided the basis for a chronological scheme that had global application (King 1967), a grand design published at a time when workers elsewhere were experiencing increasing difficulty in making compatible the numerous 'local histories' that had been identified, as was well shown in David Linton's (1964) valiant attempt to provide a synthesis of Tertiary landscape evolution in Britain.

Since the 1960s there has been increasingly widespread and severe criticism of classical denudation chronology. Some of these criticisms focused on the inadequacies of the Davisian

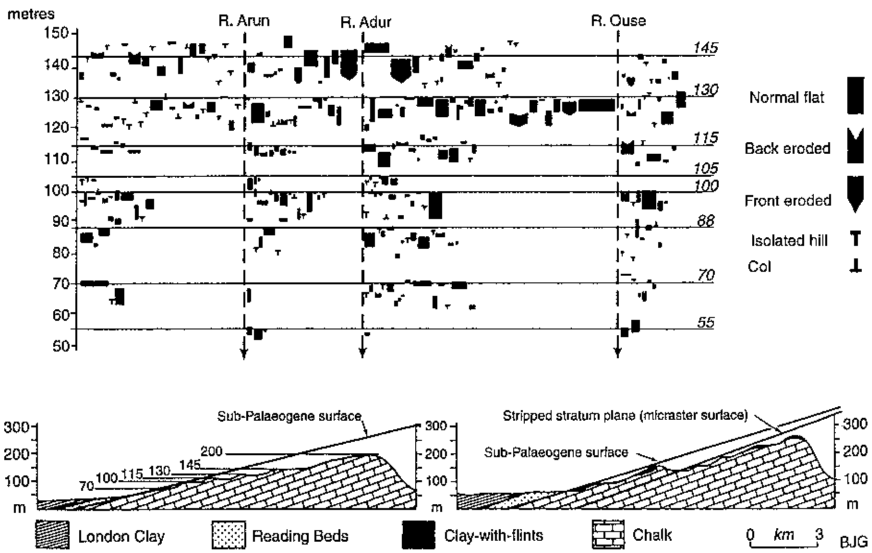

Figure 43 Height-range diagram for erosional levels identified on the South Downs backslopes by Sparks (1949), together with his interpretation of several horizontal marine benches. The two sections compare Sparks's interpretation with that of later workers (see Jones 1981)

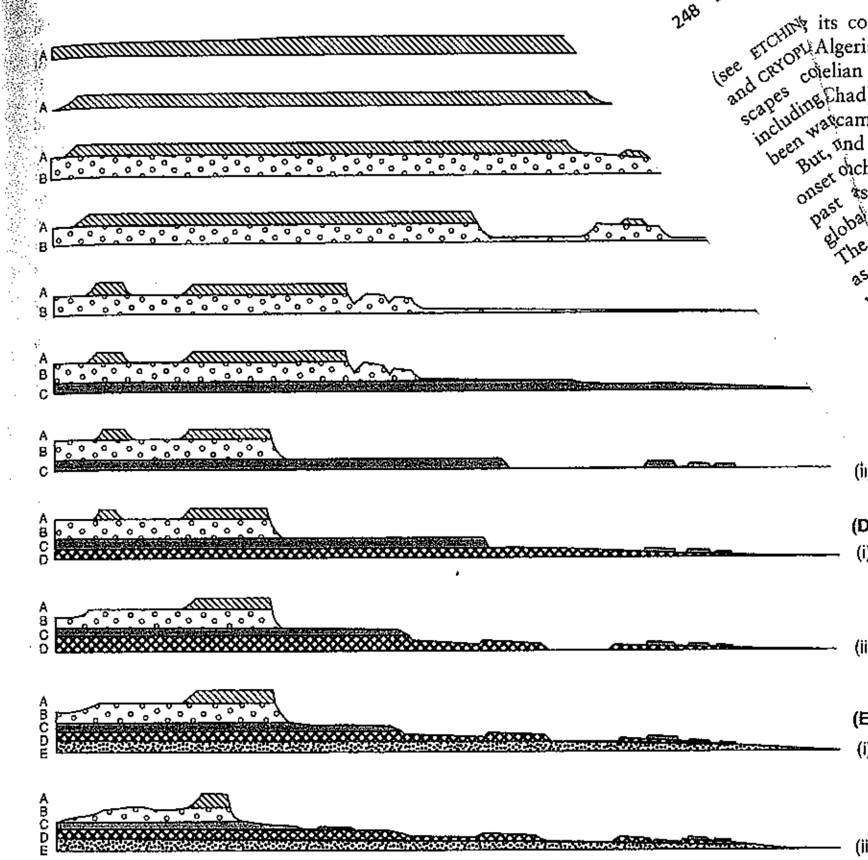

Figure 44 The topographic evolution of the area between the Drakensberg Escarpment and the Natal Coast as detailed by Lester King in 1972. Continental rupture in Jurassic–Cretaceous (A) was followed by pulses of differential uplift in the mid-Cretaceous (up to 1,250 m – B), Miocene (C), end Miocene (up to 800 m – D) and Pliocene (up to 625 m – E), each of which generated incision, backwearing and pediplanation. The result is a stepped and warped landscape in which remnants of the Mesozoic Gondwana Surface, originally at 600 m (A), survive at c.3,500 m (E(ii))

model of landscape evolution (Chorley 1965) compared with the approaches of Gilbert, Hack and Penck. Others pointed to the oversimplified theoretical concepts employed and the dangerous over reliance on morphological evidence compared with the far more rigorous, scientific approach adopted by the relatively new but rapidly expanding field of Quaternary Science, with its focus on the analysis of sediments. Yet others pointed to the difficulties of separating base level controls from structural controls and the fundamental problems of correlating surfaces over significant distances simply on the basis of elevation, especially when the origin and age of such 'surfaces' were often largely based on speculation. It came to be recognized that 'flats' or erosional levels could be produced by a wide variety of processes including pediplanation, etchplanation (see ETCHING, ETCHPLAIN AND ETCHPLANATION) and CRYOPLANATION and that contemporary landscapes could contain EXHUMED LANDFORMS, including planes of unconformity that may have been warped.

But, most importantly, the 1960s witnessed the onset of radical changes to prevailing views of the past arising from growing knowledge about global tectonics and Quaternary climate change. The new paradigms indicated 'ceaseless motion' as a characteristic of the lithosphere, together with oscillating climatic conditions and sea levels, and thereby seriously undermined notions of both the formation and survival of surfaces from the distant past, except under special conditions or where LANDSCAPE SENSITIVITY is very low.

As a result of these criticisms, denudation chronology fell into disrepute and almost became a term of abuse. Many of the detailed chronologies came to be dismissed as pure speculation or were demolished after detailed reinvestigation (as in the case of Sparks's (1949) sequence – see Figure 43). However, alternative explanations of landscape development as due to waves of aggression acting on rock sequences offering variable resistance to erosion have proved to be neither edifying nor satisfying. Most landscapes contain conspicuous morphological features (low relief surfaces, bevels, benches) that require explanation and the evolution of topographic landscapes remains a fascination. As a consequence, studies of landscape development have continued, albeit in different form and under new names. Long-term landform/landscape development or evolutionary geomorphology has many of the same aims as denudation chronology but within a much more complex conceptual framework, utilizing analysis of surficial deposits and the ever-growing range of absolute dating techniques. One of the most significant developments has been the attempt to correlate offshore sedimentary sequences with onshore denudation episodes, as pioneered by L.C. King. Some idea of the range of approaches adopted can be gained from Smith *et al.* (1999).

References

Baulig, H. (1928) *Le Plateau central de la France et sa bordure Méditerranéenne, étude morphologique*, Paris: Armand Colin.
Brown, E.H. (1960) *The Relief and Drainage of Wales*, Cardiff: University of Wales Press.
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.
Davis, W.M. (1889) The rivers and valleys of Pennsylvania, *National Geographic Magazine* 1, 183–253.
——(1909) *Geographical Essays*, Boston: Ginn and Company.
Johnson, D.W. (1931) *Stream Sculpture on the Atlantic Slope*, New York: Columbia University Press.
Jones, D.K.C. (1981) *The Geomorphology of the British Isles: Southeast and Southern England*, London: Methuen.
King, L.C. (1967) *The Morphology of the Earth*, Edinburgh: Oliver and Boyd.
——(1972) *The Natal Monocline*, Durban: University of Natal.
Linton, D.L. (1964) Tertiary landscape evolution, in J.W. Watson and J.B. Sissons (eds) *The British Isles*, 110–130, London: Nelson.
Smith, B.J., Whalley W.B. and Warke, P.A. (eds) (1999) *Uplift, Erosion and Stability: Perspectives on Long-term Landscape Development*, London: Geological Society Special Publication 162.
Sparks, B.W. (1949) The denudation chronology of the dip-slope of the South Downs, *Proceedings of the Geologists Association* 60, 165–215.
Wooldridge, S.W. and Linton, D.L. (1955) *Structure, Surface and Drainage in South-east England*, London: George Philip.

Further reading

Summerfield, M.A. (1991) *Global Geomorphology*, Ch.18, Harlow: Longman.

SEE ALSO: clay-with-flint; dating methods; drainage pattern; duricrust; dynamic equilibrium; erosion; geomorphic evolution; global geomorphology; inselberg; sea level; slope, evolution; tectonic geomorphology

DAVID K.C. JONES

DESERT GEOMORPHOLOGY

The scientific study of deserts and desert landforms had its origins in the latter half of the nineteenth century driven by many external and internal forces including imperialism, colonialism, military adventurism, romance as well as the desire to explore and exploit mineral resources and claim land for agriculture, ranching and grazing.

Some of the earliest descriptive observations on deserts can be found in the works of Greek and Roman geographers (e.g. Herodotus, Aristotle, Seneca, Strabo, Pliny, Ptolemy, among others). Arab and Muslim geographers also wrote about deserts during their various journeys within the Dar-el-Islam which stretched from Morocco to Indonesia. Deserts have also been mentioned in the writings of European travellers that went along the great Silk Road to China (via the deserts of the Middle East and Central Asia) beginning in the thirteenth century (e.g. Marco Polo and others).

The imperial ventures of Spain and Portugal in the New World during the sixteenth century led to some of the earliest descriptive observations of deserts. The Spanish, in the shape of the *conquistadores* and missionaries, were probably the first to venture in the Sonoran and Chihuahuan deserts of North America (e.g. De Vaca, Coronado, De Soto, Father Kino, among others) and provided detailed descriptions of the landforms.

Historical framework I: 1850 to 1950

With colonial aspirations and military adventurism in full swing during the middle part of the nineteenth century, England, France, Germany and the United States sent forth the first wave of soldiers, surveyors and scientists to investigate deserts which lay within their purview (Tchakerian 1995: 2).

With British administration in Egypt, some of the earliest scientific works on desert landforms were conducted from the late 1890s to the late 1920s (see the bibliographies in Goudie 1999) by Cornish, Beadnell, King and Ball, with particular emphasis on general dune forms and processes. The book *Waves of Sand and Snow* by Vaughan Cornish (1914) is one of the earliest detailed studies on bedforms and wind regime. It was Ralph Alger Bagnold (1896–1990), who, during his travels to the Western Desert in Egypt in the late 1920s, became fascinated with dune processes and dynamics and, once discharged from the British Army in 1934, built a wind tunnel in Imperial College, London to study the mechanics of blown sand and dunes. His pioneering research culminated in the now classic book *The Physics of Blown Sand and Desert Dunes* (1941), one of the most significant canons in the geomorphic literature (Goudie 1999: 6). He laid the theoretical foundations of AEOLIAN GEOMORPHOLOGY with his detailed analysis of fluid flow, particle motion, sediment transport and dune forms (particularly barchans and siefs). Bagnold's autobiography, *Wind, War and Sand* (1990), written shortly before his death, is a wonderful journey into the mind and soul of this great scientist, and, along with his earlier monograph of desert travels, *Libyan Sands* (1935), should be read by all desert scholars.

With France gradually extending its colonial empire from Morocco eastwards to Algeria and Tunisia and ultimately to all of the Sahelian countries (Mauritania, Niger, Mali and Chad), the western and central Sahara Desert became the focus of study by French scientists and their African colleagues. The work of the French geoscientists has gone relatively unnoticed outside of western Europe (most have not been translated into English) and some very notable works remain to be read and cited. An excellent synopsis is provided by Goudie (1999). A significant body of works is devoted understandably to aeolian processes and landforms (particularly to dune processes and formation, sand sea dynamics and dune orientations and wind regimes) owing to the fact that some of the world's most extensive sand seas (ergs) (see SAND SEA AND DUNEFIELD) are found in the Western Sahara. Some of the most notable contributors include Rolland, Chudeau, Aufrère (the most prolific), Capot-Rey, Dubief and Clos-Arceduc. Many of their papers were published in the *Travaux Institute de Recherches Sahariennes*, in Algiers. Another significant contributor was Emile Gautier, whose seminal monograph *Le Sahara* (1935) is one of the classic works on the general geomorphology of the Sahara Desert (including the human impact on the environment). This rich tradition of French research in the Sahara has continued in the second half of the twentieth century including works by Birot, Cailleux, Dresch and Tricart.

German colonial interests in south-west Africa (including the Namib and the Kalahari deserts) resulted in a number of expeditions to evaluate the economic potentials for the deserts of southern Africa. The works of the German geographer Passarge are especially significant including his monograph *Die Kalahari* published in 1904. In the deserts of interior Asia, German scientists such as Richthofen were one of the first westerners to recognize the significance of desert dust and LOESS. The Swedish geographer and military opportunist Sven Hedin wrote about the Gobi Desert and its landforms. The Australian arid zone was investigated by a number of explorers and scientists from the mid-nineteenth into the early twentieth centuries, driven largely by ranching and mineral resource evaluation (Cooke *et al.* 1993: 13).

The scientific study of the North American deserts begins with the United States federal exploration of lands west of the Mississippi

River (largely as a result of the acquisition of large swaths of territory beginning with the Louisiana Purchase of 1803 and ending with lands gained from Mexico as a result of the US–Mexican war of 1848). After the end of the US Civil War in 1865, the western surveys (primarily to evaluate the mineral, settlement and railroad prospects) led by King, Hayden, Powell and Wheeler produced some of the first detailed descriptions of the American deserts. John Wesley Powell's 1878 classic monograph *Report on the Lands of the Arid Region of the United States* has extensive descriptions of deserts and desert landforms and was one of the first studies to evaluate the natural resources of the region as well as its suitability for extensive human occupation (something that Powell did not recommend). G.K. Gilbert's influential *Report on the Geology of Henry Mountains* (1877) for the first time showed how process geomorphology (with its foundation in detailed fieldwork and data gathering based on principles from physics, mathematics, statistics) can contribute to the understanding of deserts and desert landforms. Gilbert's contributions to geomorphology are too numerous to cite here (he is considered the 'founder of process geomorphology') but his works on desert fluvial processes, Quaternary lakes and WIND EROSION are particularly noteworthy. In the 1920s and 1930s, Kirk Bryan published numerous influential papers on wind erosion, differential weathering and erosion, pedestal rocks, pediments, and arroyos of the southwestern USA. During the early part of the twentieth century, AEOLIAN PROCESSES were seen as the dominant sculptor of desert landforms, as promulgated by Keyes in 1912 in his 'aeoliana-tion' cycle (Tchakerian 1995: 2). William Morris Davis and his colleagues in a series of papers in the 1930s put to rest the dominating role of wind in the evolution of desert landscapes and accurately pointed out the more substantial role played by weathering, mass movement and fluvial processes. Also in the 1930s, the severe environmental consequences of the 'Dust Bowl' years in the Great Plains of the United States, led many scientists to focus their studies on wind and soil erosion, including those by W.S. Chepil and colleagues, resulting in many publications dealing with the mechanics of aeolian entrainment and transport under different land use activities (Tchakerian 1995: 3).

Historical framework II: 1950 to 2000

Global studies of deserts and desert landforms were rejuvenated in the second half of the twentieth century as a result of a number of technological and theoretical advances. In the following section, some of the major themes are briefly highlighted but detailed consideration and bibliographies in specific desert processes or landforms is left to other contributors to this volume.

Major developments were led by advances, refinement and availability of air photos (after the Second World War) and satellite imagery (during the 1970s), enabling for the first time a continental and global perspective in desert research, including the first comprehensive global dune classification using Landsat imagery (McKee 1979). The emergence of planetary geomorphology has led to more focused attention on desert landforms as terrestrial analogs for arid Mars, particularly in aeolian processes and landforms, canyon formation, sapping and other mass movement processes (e.g. Malin and Edgett 2000).

Increased studies in desert geomorphology in the latter half of the twentieth century have been driven for several reasons: (1) intrinsic fascination as distinct landforms, such as the study of desert PEDIMENTS or STONE PAVEMENTS, ALLUVIAL FANS, sand dunes. Studies in fluvial processes and landforms have been at the forefront of much desert geomorphology research during this period (e.g. Graf 1988); (2) mineral resource potential of desert landforms, such as evaporites (sulphates, nitrates, etc.) from playas, aggregates and groundwater resources in alluvial fans, uranium in DURICRUSTS; (3) desert landforms as indicators for past climatic and ecological change, such as Quaternary lakes and their sediments; (4) the archaeological value of certain desert landforms, such as the study of petroglyphs on rock coatings; and (5) the increased human occupancy and flood hazard risk on or near alluvial fans, ARROYOS, BADLANDS, as well as the environmental hazards associated with increased desert aeolian dust and mineral aerosols and their effects on human health and atmospheric visibility.

The refinement and extension of Bagnold's theoretical foundations have been the primary focus for most aeolian-related research during the past fifty years. This can be seen in the plethora of scientific papers, monographs and symposia proceedings devoted exclusively to aeolian processes and landforms during the past two decades alone (Tchakerian 1995: 4). Associated with the above refinements were developments in mathematical modelling, computer simulations, and the use of complex, non-linear dynamical systems theory for understanding wind flow and bedforms (e.g. Werner and Kocurek 1999). The resurgence of aeolian geomorphology has also been stimulated by the fact that aeolian sedimentary environments are good analogues for studying hydrocarbon reservoirs in the geologic record. Advances in the understanding of single dune formation and dynamics has now led researchers to tackle the more daunting task of analysing draas (megadunes) and ergs (e.g. Pease and Tchakerian 2002). Other significant developments include the application of luminescence techniques for dating aeolian sands and Quaternary dune systems, and the wide availability of very sophisticated instruments (anemometers, electron microscope, sediment traps) and data loggers for the gathering and analysis of wind data and sediments. Technological advances were also instrumental in heralding numerous studies in aeolian desert dust and mineral aerosols culminating in a series of papers assessing the global distribution of major atmospheric dust sources (e.g. Prospero *et al.* 2002). Research in aeolian processes and landforms has also been driven by concerns for land degradation (DESERTIFICATION) and global environmental change.

The growing scientific interest in desert geomorphology of the past three decades has led to the establishment of many centres of desert research, new journals, international associations, and to UN-sponsored research and publications. This has led to vigorous exchanges of ideas and co-operation among Earth scientists unconstrained by traditional disciplinary boundaries.

References

Cooke, R.U., Warren, A. and Goudie, A.S. (1993) *Desert Geomorphology*, London: UCL Press.
Goudie, A.S. (1999) The history of desert dune studies over the last 100 years', in A.S. Goudie, I. Livingstone and S. Stokes (eds) *Aeolian Environments, Sediments and Landforms*, 1–13, London: Wiley.
Graf, W. (1988) *Fluvial Processes in Dryland Rivers*, Berlin: Springer-Verlag.
McKee, E.D. (ed.) (1979) *A Study of Global Sand Seas*, United States Geological Survey Professional Paper No. 1,052.
Malin, M.C and Edgett, K.S. (2000) Evidence for recent groundwater seepage and surface runoff on Mars, *Science* 288, 2,330–2,335.
Pease, P.P. and Tchakerian, V.P. (2002) Composition and sources of sand in the Wahiba Sand Sea, Sultanate of Oman, *Annals of the Association of American Geographers* 92, 416–434.
Prospero, J.M., Ginoux, P., Torres, O., Nicholson, S.E. and Gill, T.E. (2002) Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus-7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product, *Reviews of Geophysics* 40, 1–29.
Tchakerian, V.P. (1995) The resurgence of aeolian geomorphology, in V.P. Tchakerian (ed.) *Desert Aeolian Processes*, London: Chapman and Hall.
Werner, B.T. and Kocurek, G. (1999) Bedform spacing from defect dynamics, *Geology* 27, 727–730.

Further reading

Thomas, D.S.G. (ed.) (1997) *Arid Zone Geomorphology*, London: Belhaven.

VATCHE P. TCHAKERIAN

DESERT VARNISH

Desert varnish, a paper-thin deposit, drastically darkens the appearance of desert rocks. Any rock type can host desert varnish, so long as its surface remains stable for the thousands of years it usually takes varnish to accrete. Rock varnish is the preferred term where this ROCK COATING occurs in non-desert settings, for example, alpine, Antarctic, Arctic, periglacial, stream, temperate and tropical environments. The term desert varnish is most often used in arid regions.

Like other rock coatings, desert varnish is deposited on rock surfaces and does not derive from the host rock itself. Arrows in the middle image in the middle row of Plate 34 exemplify this discrete contact. Like CASE HARDENING and other rock coatings, many varnishes seen at the surface today actually start in the subsurface in fissures. Varnishes are usually less than 100 μm thick, and even where micro-basins host deposits of a few hundred micrometres, median thicknesses are usually less than 10 μm thick.

Wind does not cause shiny varnish; wind abrades away this relatively soft coating. In fact, the presence or absence of desert varnish is an important clue that a particular desert pavement was not or was made by aeolian deflation. Usually dull in lustre, its occasional sheen comes from a smooth surface micromorphology in combination with manganese enrichment at the very surface of the varnish.

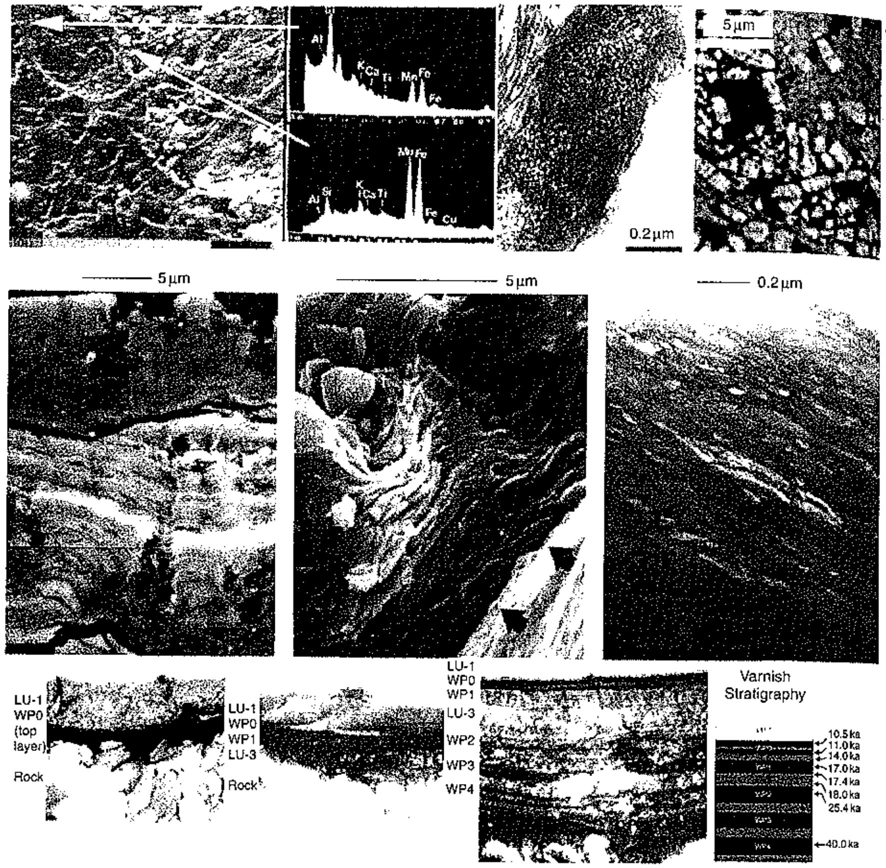

Plate 34 Microscopic views of desert varnish from arid environments. Top row: microscopic evidence for bacterial origin of rock varnish from left to right: secondary electron image of Negev Desert budding bacteria where the bacteria greatly enhanced manganese and iron; transmission electron image of manganese encrusting a bacterial form; and backscattered electron image of bacteria revealed by acid etching. Middle row: layering of desert varnish shown in backscattered (left), secondary (middle), and high resolution transmission electron microscope (right) images. Bottom row: calibration of microlaminations seen in optical microscopic views of ultra-thin sections, where black layers in the varnish represent wet periods that have been calibrated by Tanzhuo Liu's research (Liu *et al.* 2000). The thin sections from left to right show progressively older varnishes with progressively more complex layers from Death Valley, California

Clay minerals are the major ingredient of desert varnish, typically comprising more than half and sometimes as much as 90 per cent. The clay minerals impose the layered structure seen in the middle row of Plate 34. Clays are deposited as dust on rock surfaces, and are then cemented to the host rock by hydroxides and oxides (Potter and Rossman 1979) of manganese (birnessite) and iron (goethite and hematite). Manganese and iron make up about a third of varnish, with

typically less than 5 per cent of varnish composed of other components.

The mystery of desert varnish surrounds how to explain the great abundance of manganese, the element that gives varnish its dark brown to black colour. The elemental abundance of manganese in varnish is 50 to 300 times the concentrations found in dust that falls on rock surfaces. Put another way, ratios of manganese to iron are about 1:40 to 1:60 in surrounding soils and dust, but are about 1:1 in varnish.

In the past century, there have been two competing models to explain manganese enrichment. The first was a chemical process favoured by geochemists, whereby naturally acidic rain dissolves manganese in the rock or dust (but not the iron). Then, manganese oxidizes upon exposure to a slightly higher pH. The competing model was a microbial process, whereby bacteria precipitate manganese (Drake *et al.* 1993).

Although bacteria have been cultured from varnish and have made 'artificial varnish' in the laboratory (Dorn 1998), the typically slow rate of varnish growth (on average, about a micrometre per thousand years) makes it very difficult to have confidence that bacteria cultured today in the laboratory make varnish. In fact, the type of gram-positive bacteria most easily cultured from desert varnish today have not yet been identified within varnish layers. To make the matter more difficult, 'biomolecular fossils', such as amino acids generated by these bacteria, exist in both desert varnish and unvarnished weathering materials. Thus, most convincing evidence for a bacterial mechanism is actually seeing manganese enhancement *in situ*. In the upper row of Plate 34, budding bacteria can be seen concentrating manganese and iron.

New high resolution transmission electron microscope evidence (Krinsley 1998) reveals that these chemical and biological models are not truly in competition, but work in tandem. Varnish formation can be explained by a four-step process. Step 1 is the enhancement of varnish (and to a lesser extent iron) by bacteria; the top row in Plate 34 shows manganese-rich sheaths of bacteria. Step 2 is the chemical dissolution of the bacterial sheaths, whereby manganese and iron are broken down into nanometre-sized granules. Step 3 is chemical transport of manganese and iron into clay minerals. Step 4 is the precipitation of unit cells of manganese and iron inside clay minerals. Potter and Rossman (1979) noted that the hexagonal arrangement of oxygens in clay mineral layers form a template for crystallization of the manganese mineral birnessite seen in desert varnish.

Krinsley (1998) shows high resolution imagery revealing all steps in this polygenetic process whereby clay minerals and oxides are co-dependent in varnish formation. Clay provides the overall structure and template for oxide precipitation, while bacteria simply provide a ready source of manganese (and iron) cement. Varnish formation all takes place within a few micrometres of the bacterial source, where the manganese and iron are redistributed with hygroscopic water – all inside layers like those seen in the middle row of Plate 34.

Environmental changes play an important role in the development of desert varnish. Where lichens start to grow, for example, biological acids destroy desert varnish by dissolving the manganese and iron oxides. Where rocks come to exist in a desert pavement, a ground-line band of very thin and shiny varnish forms a circle around a desert pavement clast. But where varnish grows on the tops of boulders less influenced by local environmental changes, regional climatic change plays an important role in varnish formation.

In these boulder-top varnishes wetter climates favour bacterial enhancement, yielding layers that are particularly rich in manganese. Drier climates with more alkaline dust produce layers that are not as rich in manganese. The bottom row in Plate 34 shows desert varnishes from Death Valley, California, where growth of these layers has been calibrated using a combination of numerical dating methods (Liu *et al.* 2000; Zhou *et al.* 2000). Progressively older varnishes show progressively more complex layers. Varnish microlaminations provide archaeologists and geomorphologists with a powerful tool, because they reveal both climatic change and a time signal.

Some of the most interesting aspects of desert varnish surround its minor and trace constituents. Lead, for example, is greatly enhanced in the uppermost micron of the varnish from twentieth-century air pollution. The carbon that is trapped within and underneath varnish shows some potential as a means of radiocarbon dating varnish, but the history of the carbon is usually too complicated to make this technique useful. Mobile trace elements decline progressively over time, as they are leached by hygroscopic and capillary water (Krinsley 1998). Varnish also traps foreign material crushed into rock engravings as a part of religious ceremonies (Whitley *et al.* 1999). New experimental studies of trace isotopes such as ^{7}Be, $\delta^{13}C$ and $\delta^{17}O$ show potential to reveal new insights into this ubiquitous weathering phenomenon. Some planetary scientists believe that desert varnish exists on Mars, and that Martian varnish might preserve active organisms or at least biological fossils such as those seen in the top row of Plate 34.

References

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

Drake, N.A., Heydeman, M.T. and White, K.H. (1993) Distribution and formation of rock varnish in southern Tunisia, *Earth Surface Processes and Landforms* 18, 31–41.

Krinsley, D. (1998) Models of rock varnish formation constrained by high resolution transmission electron microscopy, *Sedimentology* 45, 711–725.

Liu, T., Broecker, W.S., Bell, J.W. and Mandeville, C.W. (2000) Terminal Pleistocene wet event recorded in rock varnish from the Las Vegas Valley, southern Nevada, *Palaeogeography, Palaeoclimatology, Palaeoecology* 161, 423–433.

Potter, R.M. and Rossman, G.R. (1979) The manganese- and iron-oxide mineralogy of desert varnish, *Chemical Geology* 25, 79–94.

Whitley, D.S., Dorn, R.I., Simon, J.M., Rechtman, R. and Whitley, T.K. (1999) Sally's Rockshelter and the archaeology of the vision quest, *Cambridge Archaeological Journal* 9, 221–247.

Zhou, B.G., Liu, T. and Zhang, Y.M. (2000) Rock varnish microlaminations from northern Tianshan, Xinjiang and their paleoclimatic implications, *Chinese Science Bulletin* 45, 372–376.

SEE ALSO: rock coating

RONALD I. DORN

DESERTIFICATION

Desertification is a term for land degradation caused by adverse human impacts in arid, semi-arid and dry-subhumid areas, together called the 'susceptible drylands' (Middleton and Thomas 1997). Hyper arid areas are in general terms not regarded as sites of desertification, since these areas are extremely desert-like due to natural conditions. The term desertification has also been used more widely, for example to refer to land degradation in non-dryland contexts, and there are over a hundred published definitions. There is widespread consensus that the term should be restricted to susceptible drylands, a view embodied in the 1994 UN Convention to Combat Desertification (UNCCD), which by March 2002 had been ratified by the governments of 179 countries. Desertification occurs not just in the developing world, but also under the impact of the inappropriate or excessive use of agricultural technologies in the dryland areas of the developed world. Overall, desertification relates to the unsustainable use of the land in drylands.

The term desertification is widely regarded to have been coined by Aubreville (1949), who used it to describe the environmental impact of forest clearance in West Africa, leading in his opinion to the creation of an ecological desert. It gained renewed use in the 1960s and 1970s (and was sometimes used interchangeably with desertization) when the major Sahel drought at the time led to a decline in biomass, famine and livestock and human deaths in the Sahel zone along the southern margin of the Sahara Desert.

Desertification has both social and environmental components. Social dimensions relate to the pressures and processes that cause people to carry out activities that degrade the land. Today, international efforts to reduce desertification, such as the UNCCD, lay great emphasis, particularly in the developing world, on addressing the social and economic issues that lead people to carry out land use practices that lead to land degradation. Environmental dimensions of desertification are important too, however, and relate to the actual physical processes of land degradation and to distinguishing impacts of natural dryland environmental variability from anthropogenically aggravated changes to land systems. The latter is important since it is possible to confuse the impacts of dryland rainfall variability with changes in soils and vegetation caused by human actions (Thomas and Middleton 1994). These confusions have led, in the past, to overestimates of the physical dimensions and scale of desertification, including for example, inappropriate statements being made about the advance of deserts over productive land (see Hellden 1991; Thomas 1993). The causes and use of the term desertification are complex, however, since natural drought may itself place pressures on social and agricultural systems that in turn lead to land degradation.

Physical processes of desertification

The physical processes of desertification comprise depletion of and damage to the soil, ground water, and to some extent vegetation systems, that reduces their productive capacity or biological potential. The soil provides the geomorphological context of desertification, with loss of productive potential occurring either through the physical loss of soil by water or wind erosion, or internal physical and chemical changes such as compaction, salinization, alkalinization and nutrient depletion. Dryland soils may be thin or skeletal, with generally slow rates of formation due to the limited availability of water for the weathering of underlying bedrock and also to slow build up of organic material. Exceptions may occur in specific geomorphic locations such as valley floors where seasonal water supply may be greater. Natural recovery from soil erosion or internal changes is therefore likely to be slow.

There has been considerable debate regarding whether changes in dryland vegetation systems that are unaccompanied by soil system changes constitute desertification. These debates have arisen for a number of reasons: many savanna systems are 'non-equilibrium' ecosystems that do not achieve a spatial or temporal steady state because of the natural dynamism of dryland climates (Mistry 2000); dryland vegetation systems can be both resilient and adapted to disturbance and can exhibit rapid recoveries; the impacts of droughts and land degradation on vegetation systems can be hard to distinguish; observed vegetation changes caused by human pressures may not be accompanied by soil system changes (Dougill *et al.* 1999), facilitating recovery if land use pressures are reduced or removed. It should be noted however that vegetation depletion can set the scene for desertification processes to take effect via the processes of erosion.

Geomorphological dimensions of desertification

Desertification via erosion processes may be relatively easy to recognize and in susceptible drylands both wind and water erosion can be important. WIND EROSION potential is greatest in areas of low relief and unconsolidated sediments, for example the Canadian Prairies and the midwest of the USA, as witnessed by the occurrence of severe DUST STORMS during the 1930s, while areas of steeper topography are more susceptible to water erosion, for example in the Highlands of Ethiopia and Kenya and around the Mediterranean basin in Europe and North Africa.

Changes within the soil due to human activities are, with the possible exception of salinization, less visible and more insidious than those caused by erosion. Salinization and alkalinization associated with irrigation schemes are widely cited causes of productivity decline in drylands. Other internal changes relate to waterlogging, also associated with irrigation, and the crusting and compaction of soils, increasingly caused by the mechanized cultivation procedures used in agriculture. Nutrient depletion is a notably widespread but often underestimated form of desertification (Thomas and Middleton 1994). Nutrient loss can be caused by the actual physical removal of soil by erosion but is often a function of the clearance of natural vegetation to create fields for cultivation, the subsequent intensity of cultivation or the lack of application of fertilizers, especially in developing world dryland areas.

Assessing the extent and nature of desertification

There are no readily available means of gaining absolute data on the occurrence of desertification. Earth observation via remote sensing can be useful for detecting dimensions of vegetation change, and for distinguishing natural fluctuations due to rainfall variability from changes caused by human actions (e.g. Tucker *et al.* 1991), but even the highest resolution imagery can be too coarse to identify many elements of soil erosion and unsuitable for determining soil internal changes. Thus estimates of the global extent of desertification are likely to be crude, and sometimes highly erroneous (Thomas 1993). Field studies and modelling approaches are important for local and regional investigations of degradation (Mairota *et al.* 1998). Increasingly, as the social dimensions of land degradation are recognized as vital elements of any efforts to ameliorate the problem, land user understanding and knowledge, often untapped, especially in the developing world, are viewed as essential for an effective understanding of where, why and how desertification occurs (Reed and Dougill 2002).

The UN has been the instigator of the few attempts to assess the global scale of desertification. Estimates produced in the 1970s and 1980s have subsequently been heavily criticized in the scientific community for their lack of methodological rigour and for confusing natural cyclic changes in vegetation systems due to drought impacts with desertification caused by human actions.

The most recent global assessment of land degradation caused by soil degradation (GLASOD) was carried out by the International Soil Reference Centre in the Netherlands on behalf of UNEP in the late 1980s and early 1990s. A GIS was used to analyse data collected through a clearly defined, but somewhat qualitative, methodology (Middleton and Thomas 1997). Despite its flaws, GLASOD has provided a database through which assessments of susceptible dryland soil degradation can be analysed in terms of geographical coverage (see Table 10), contributory degradation processes, and relationships to land use.

GLASOD estimates that in the late 1980s and early 1990s approximately 1,030 million hectares, equivalent to 20 per cent of the susceptible drylands, had experienced soil degradation processes caused by human activities. Water erosion was identified as the major physical process of degradation in 48 per cent of this area and wind erosion in 39 per cent. Chemical degradation (mainly salinization, alkalinization and nutrient depletion) was dominant in only 10 per cent of the area, and physical changes such as compaction and crusting in just 4 per cent. These latter figures may well be underestimated given the problems of identification of these less visible processes. The severity of degradation was described by GLASOD as strong or extreme in 4 per cent of the susceptible drylands – meaning that land where the original biotic functions of the soil have been destroyed, and which are irreclaimable without major restorative measures does not occur widely.

What human actions lead to desertification?

Almost any land use in the susceptible drylands has the potential to lead to desertification, if it is conducted to excess or in locations to which it is not well suited. The literature on desertification widely cites four forms of land use as major contributors to the problem: cultivation, irrigation schemes, livestock production and deforestation.

Overcultivation has sometimes been viewed as the main cause of desertification, especially in areas of the developing world where increasing populations have necessitated attempts to increase yields without the resources available for additional fertilizers. Nutrient depletion is therefore a potentially serious issue, with the World Bank attributing declining yields of staple crops in Sahel nations and in parts of South America to this problem. Attempts to increase crop yields through mechanization can lead to soil compaction, increasing runoff and erosion under intensive dryland rainfall events. Aeolian deflation can also be exacerbated by the removal of shelter belts to allow large machinery to be used.

Irrigation systems, whether by canals from storage dams or through centre pivot systems, can cause desertification through waterlogging, salinization and alkalinization, the excessive accumulation of sodium in the soil. High evapotranspiration rates in drylands mean that excessive irrigation can lead readily to the accumulation of salts in the soil. Waterlogging arises when irrigation rates are so high that the water table is raised excessively.

Table 10 Dryland soil degradation (million ha) by continent according to GLASOD

	Susceptible dryland area	Light and moderate desertification	Strong and extreme desertification	Total desertified
Africa	1,286.0	245.3	74.0	319.3
Asia	1,671.8	326.7	43.7	370.4
Australasia	663.3	86.0	1.6	87.6
Europe	299.7	94.6	4.9	99.5
North America	732.4	72.2	7.1	79.3
South America	516.0	72.8	6.3	79.1
Total	5,169.2	897.6	133.7	1,035.2

Source: Derived from data in Middleton and Thomas (1997)

Many dryland regions may be better suited to extensive livestock production than to cultivation. Pastoralism is seen as a contributor to desertification when it is over intensive, which may occur in developed world drylands and in the developing world when traditional practices are replaced by commercial systems. What may result is excessive grazing that can alter plant community composition and reduce overall plant cover, thereby increasing the susceptibility of the land to erosion processes. The role of pastoralism in desertification is however complex and somewhat controversial: as noted earlier vegetation changes may not always be accompanied by soil system changes and therefore may be reversible.

Deforestation in drylands is associated both with the clearance of lands for cultivation and, in developing world contexts, with the collection of wood for use as the dominant domestic fuel. The greatest threat deforestation offers for desertification is in areas of steep slopes. For example, during the twentieth century a tenfold reduction in woodland cover occurred in the Ethiopian Highlands through the effects of expanding cultivation and fuelwood collection: it is not surprising that soil erosion is a major problem in this area.

References

Aubreville, A. (1949) *Climats, forêts et desertification de l'Afrique tropicale*, Paris: Editions géographiques maritimes et colonials.

Dougill, A.J., Thomas, D.S.G. and Heathwaite, A.L. (1999) Environmental change in the Kalahari: integrated land degradation studies for non equilibrium dryland environments, *Annals, Association of American Geographers* 89, 420–422.

Hellden, U. (1991) Desertification: time for a reassessment? *Ambio* 20, 372–383.

Mairota, P., Thornes, J.B. and Geeson, N. (1998) *Atlas of Mediterranean Environments in Europe: The Desertification Context*, Chichester: Wiley.

Middleton, N.J. and Thomas, D.S.G. (1997) *World Atlas of Desertification*, London: UNEP/Edward Arnold.

Mistry, J. (2000) *World Savannas: Ecology and Human Use*, Harlow: Prentice Hall.

Reed, M. and Dougill, A.J. (2002) *Geographical Journal*, 163, 195–210.

Thomas, D.S.G. (1993) Sandstorm in a teacup? Understanding desertification in the 1990s, *Geographical Journal* 159, 318–331.

Thomas, D.S.G. and Middleton, N.J. (1994) *Desertification: Exploding the Myth*, Chichester: Wiley.

Tucker, C.J., Dregne, H.E. and Newcomb, W.W. (1991) Expansion and contraction of the Sahara Desert, 1980 to 1990, *Science* 253, 299–301.

DAVID S.G. THOMAS

DESICCATION CRACKS AND POLYGONS

As evaporation of water from a saturated, fine-grained, cohesive sediment occurs, volume reduction may be accompanied by sufficient tensional stress for rupture to take place. Cracks are thereby formed and may display polygonal patterns. The morphology of the rupture patterns depends both on material properties (structure, degree of packing, moisture content, etc.) and on environmental conditions (temperature, humidity, rate of drying, etc.) (Corte and Higashi 1964). Cracks and polygons of this type are common geomorphic phenomena of both periglacial and arid environments (Maizels 1981), of vertisols and of drained proglacial lakes (Huddart and Bennett 2000).

Cracks tend to be fairly straight or smoothly curved in plan. However, their patterns, lengths, depths, widths and number show great variation. The plan form of blocks between cracks, which is determined by the crack pattern, can be flat, convex, concave or irregular, but the size of blocks tends to increase with the thickness of the material of which they are composed. The number of cracks is generally inversely proportional to sediment size, being greatest in materials rich in silt and clay. Mean crack length is proportional to sediment moisture content. It tends to decrease with time, because new, short cracks continue to be formed during each new cycle of drying. The spacing of cracks may increase with the rate of desiccation and with the proportion of clay present and according to the nature of the clay type. Sediment that is montmorillonite rich, for example, is prone to greater contraction than one with a comparable proportion of kaolinite. The presence of stones in and on fine-grained sediments may also affect the nature of the cracking.

Lachenbruch (1962) identified two common crack patterns. One is an orthogonal pattern in which cracks meet at right angles. The other is a non-orthogonal pattern characterized by triradial intersections that form obtuse angles of around 120°. The former are probably characteristic of inhomogeneous or plastic media in which stress accumulates gradually. Cracks form first at loci of low strength or high stress concentration. Because the cracks do not form simultaneously, a new crack has a tendency to join a pre-existing one orthogonally. The latter systems form in more homogeneous, relatively non-plastic media which are dried uniformly.

Particularly large 'giant desiccation cracks' are especially common on salty, playa surfaces (Neal *et al.* 1968) and may result not only from desiccation, but also because of such factors as salt mobilization, subsidence caused by groundwater withdrawal and seismic activity. The sudden creation of giant fissures can damage engineering structures (Al-Harthi and Bankher 1999; Corwin *et al.* 1991).

References

Al-Harthi, A.A. and Bankher, K.A. (1999) Collapsing loess-like soil in western Saudi Arabia, *Journal of Arid Environments* 41, 381–399.

Corte, A.E. and Higashi, A. (1964) *Experimental Research on Desiccation Cracks in Soil*. US Army Material Command Cold Regions Research and Engineering Laboratory Research Report 66.

Corwin, E.J. Alhadeff, S.C., Oggel, S.P. and Shlemon, R.J. (1991) Earth fissures, urbanisation and litigation: a case study from the Temecula area, southwestern Riverside County, California, *IAHS Publication* 200, 291.

Huddart, D. and Bennett, M.R. (2000) Subsidence structures associated with subaerial desiccation-crack piping and their role in drainage evolution on a drained proglacial lake bed: Hagavant, Iceland, *Journal of Sedimentary Research Section A: Sedimentary Petrology and Processes* 70, 985–993.

Lachenbruch, A.H. (1962) *Mechanisms of Thermal Contraction Cracks and Ice-wedge Polygons in Permafrost*, Geological Society of America Special Paper 70.

Maizels, J.K. (1981) *Freeze–thaw Experiments in the Simulation of Sediment Cracking Patterns*, Bedford College, London, Papers in Geography 13.

Neal, J.T., Langer, A.M. and Kerr, P.G. (1968) Giant desiccation polygons of Great Basin playas, *Geological Society of America Bulletin* 79, 69–90.

A.S. GOUDIE

DEW POND

Dew ponds are closed depressions, often filled with water, usually associated with chalk downland in the southern UK. The constant supply of water to these depressions on permeable substrates and frequently close to the top of slopes has caused considerable discussion over the past two centuries. As the name indicates, some writers have suggested dew (or mist or fog) as the source of replenishment but others believe that rainfall and surface runoff provide sufficient supply. Many dew ponds were dug on agricultural land, occasionally by people employed specifically for this purpose who moved from farm to farm, and the ponds were lined with clay, straw and more recently cement. Today some are dry because of lack of maintenance of the lining. There is considerable confusion in assigning origins to closed depressions in calcareous areas (see KARST) because some are natural DOLINES or sink holes while others have anthropogenic origins as dew ponds or as pits dug for chalk, marl or clay. Some even have complex origins, forming initially as a doline but subsequently being lined and used as a dew pond.

Further reading

Prince, H.C. (1964) The origin of pits and depressions in Norfolk, *Geography* 49, 15–32.

Pugsley, A.J. (1939) *Dewponds in Fable and Fact: A Collection and Criticism of Exciting Knowledge on these Curiosities*, London: Country Life.

Snow, M. (2002) *Dewponds*. http://www.dewponds.info/(Accessed 18th September 2002).

IAN LIVINGSTONE

DIAGENESIS

All the changes (physical, chemical and biological) undergone by a sediment after its initial deposition, exclusive of weathering and metamorphism (and incorporates processes such as reworking, authigenesis, replacement, leaching, hydration, bacterial action, and the formation of concretions). The term was coined by Gümbel in 1868 (*diagenese*), though no universal definition exists.

Reference

Gümbel, C.W. von (1868) *Geognostische Beschreibung des ostbayerischen Grenzgebirges oder des bayerischen und oberpfalzer Waldgebirges*, Gotha: Justus Perthes; Bavaria: Bayerisches schreibung des Koenigreichs Bayern, v. 2.

Further reading

Larson, G. and Chilingar, G.V. (eds) (1979) *Diagenesis in Sediments and Sedimentary Rocks. Developments in Sedimentology* 25A, Oxford: Elsevier Scientific.

SEE ALSO: lithification

STEVE WARD

DIAMICTITE

Diamictite (paraconglomerate, mixtite, diamixtite) is a non-genetic term for sedimentary rock consisting of sand and/or larger particles resting in a muddy matrix (Flint *et al.* 1960). Its unlithified counterpart is known as diamicton, and diamict is a general term comprising both consolidated and non-consolidated deposits.

Diamictites consist of a wide range of particles with matrix dominating, giving an overall appearance of clasts chaotically dispersed through structureless or laminated mud. Clasts are dropstones sporadically deposited on the soft substrate due to ice rafting or volcanic explosions. They are angular to subrounded, may be slightly imbricated (see IMBRICATION) and often come from remote sources hundreds of kilometres away. If lamination is present, drapes (see DRAPE, SILT AND MUD) around clasts caused by post-depositional compaction (see COMPACTION OF SOIL) occur frequently.

The origin may be by GLACIERS, DEBRIS FLOWS or turbidity currents. Most diamictites are interpreted as lithified tills (tillites) or GLACIMARINE deposits of pre-Quaternary glaciations (Schermerhorn 1974; Hambrey and Harland 1981) because of a range of features resembling tills known from modern glacial environments, such as lack of sorting, grain sizes from clays to blocks, clasts bearing GLACIAL EROSION features (e.g. STRIATIONS and polishing), glacidynamic structures (e.g. shear planes (see SHEAR AND SHEAR SURFACE) and folds), and mixed lithological components corresponding to ice flow paths. Glacial origin of most diamictites is also supported by intimate association with striated bedrock surfaces, varved clays and lithified GLACIFLUVIAL deposits, arctic fauna, periglacial (see PERIGLACIAL GEOMORPHOLOGY) structures and the correspondence to polar regions of the past.

Diamictites of glacial and glacimarine origin are known from many regions of the Earth (Miller 1996: 469–483). The Late Palaeozoic Dwyka Formation occupies extensive areas in southern Africa and consists of facies deposited mainly under disintegrating ice shelves, near the grounding line and in FJORDs by tidewater glaciers (Visser 1991). Also the Lower Proterozoic Gowganda Formation, which covers more than 12,000 km^2 on the Canadian Shield is possibly a glacimarine diamictite (Eyles *et al.* 1985). Terrestrial tillites are known from the Upper Ordovician of north-west Africa (6–8 million km^2), Upper Proterozoic of western Mauritania and north Norway, Permo-Carboniferous of the Transantarctic Mountains, and Upper Palaeozoic of Oman and Brazil.

References

Eyles, C.H., Eyles, N. and Miall, A.D. (1985) Models of glaciomarine sedimentation and their application to the interpretation of ancient glacial sequences, *Palaeogeography, Palaeoclimatology, Palaeoecology* 51, 15–84.

Flint, R.F., Sanders, J.E. and Rodgers, J. (1960) Diamictite, a substitute term for symmictite, *Geological Society of America Bulletin* 71, 1,809.

Hambrey, M.J. and Harland, W.B. (eds) (1981) *Earth's Pre-Pleistocene Glacial Record*, Cambridge: Cambridge University Press.

Miller, J.M.G. (1996) Glacial Sediments, in H.G. Reading (ed.) *Sedimentary Environments: Processes, Facies and Stratigraphy*, 454–484, Oxford: Blackwell Science.

Schermerhorn, L.J.G. (1974) Late Precambrian mixtites: Glacial and/or nonglacial? *American Journal of Science* 274, 673–824.

Visser, J.N.J. (1991) The paleoclimatic setting of the late Paleozoic marine ice sheet in the Karoo Basin of Southern Africa, in J.B. Anderson and G.M. Ashley (eds) *Glacial* Marine Sedimentation: Paleoclimatic Significance, Geological Society of America Special Paper 261, 181–189.

Further reading

Prothero, D.R. and Schwab, F. (1996) *Sedimentary Geology*, New York: W.H. Freeman.

JAN A. PIOTROWSKI

DIAPIR

Vertical intrusions, bulbous or cylindrical in shape, resulting from the upward movement of mobile materials, such as salt (halite), magma, mud and ice, which lie beneath more competent strata (see MUD VOLCANO).

Further reading

Vendeville, B.C., Hart, Y. and Vigneresse, J.L. (eds) (2000) *Salt, Shale and Igneous Diapirs in and around Europe*, Special Publication of the Geological Society of London 174.

A.S. GOUDIE

DIASTROPHISM

A general term for the various types of tectonic processes that change the level, position and altitude of the Earth's surface. It is derived from the Greek word *diastrophos*, which means 'turned', 'twisted' or 'distorted'. There are five classes of diastrophic movement (Chorley *et al.* 1984: 98): (1) orogenesis; (2) epeirogeny; (3) isostasy; (4) igneous (including volcanic); and (5) eustasy.

Reference

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

A.S. GOUDIE

DIATREME

Vents and pipes which have been injected through sedimentary strata by the explosive release of magmatic gases and filled with the products of the eruption, and fragments torn from the side of the pipes. Kimberlite pipes are examples of diatremes. Some maars, such as those in Germany, are lakes that are the surface expression of diatremes. They can be rich in environmental information (Narcisi 1996).

Reference

Narcisi, B. (1996) Tephrochronology of a late Quaternary lacustrine record from the Monticchio Maar (Vulture Volcano, southern Italy), *Quaternary Science Reviews* 15, 155–165.

A.S. GOUDIE

DIGITAL ELEVATION MODEL

A numerical description of the ground surface is helpful in addressing many geomorphological problems. Continuous topography may be quantified by a digital elevation model (DEM), any spatial array of terrain heights but most commonly a square mesh, or grid. DEM point spacing, or horizontal resolution, varies from fine (1 m) to coarse (≥5 km), depending on the application, required level of detail, and the limitations of computer storage and processing (Table 11).

DEM-related nomenclature can be inconsistent. Digital terrain model (DTM), a frequent synonym for DEM, also is applied loosely to any calculated result, such as a map of slope gradient, rather than to the input heights themselves. Digital terrain *modelling*, confusingly also DTM, widely denotes DEM processing or GEOMORPHOMETRY.

For efficient computation and display, terrain heights are arranged in a defined structure (Figure 45), usually a regular grid, a triangulated irregular network (TIN), or digitized height contours or slope lines normal to contours. DEM structures are compromises, each with advantages and drawbacks.

Square-grid, actually rectangular, DEMs (hexagons or equilateral triangles are rare) store height Z as an array of implicit longitude X and latitude Y co-ordinates (Figure 45A). While this regular and discretized (discontinuous) structure is ill matched to the varied intricacies of surface features, a grid optimizes algorithm development, data processing and registration with spacecraft images. Grid DEMs can adapt somewhat to complex terrain by recursively subdividing squares, but computational efficiency declines.

The irregularly distributed heights of a TIN (Figure 45B) are vertices of triangles that vary in shape and size. A TIN is interpolated directly from surveyed points or discrete features that are extracted manually from maps or by computer from a grid or contour DEM. The TIN is adaptive, or surface specific: it aligns with ridges and channels and has many heights in complex areas but few redundant heights in planar terrain. Offsetting these advantages is the storage and processing burden imposed by the explicit X,Y co-ordinates required for each value of Z.

Ground-surface form is neither rectilinear nor triangular. Although not all topography is fluvial, the DEM structure best reflecting processes of erosion and deposition mimics paths of steepest gradient (Moore 1991). Terrain heights at intersecting contours and interpolated slope lines (Figure 45C) comprise an adaptive DEM that defines quadrilateral units of varied size and shape – most significantly the hillside concavities followed by surface flow. Explicit X,Y co-ordinates are necessary.

Early DEMs were created by field survey, manual interpolation of topographical maps, or semi-automated tracing of contours coupled with computer interpolation. Photogrammetric profiling and later optical scanning and automated interpolation of contours replaced these techniques. Grid DEMs now are available for the Earth, its seafloor, and the planet Mars. GTOPO30, compiled from many contour maps from several sources (Gesch *et al.* 1999), covers Earth at 30′ (nominally 1 km) resolution; two older global DEMs, ETOPO5 and TerrainBase,

Table 11 Sources and applications of Digital Elevation Models (DEMs)

DEM spacing	Sources of height measurements	Some geomorphological applications
1–50 m	Contours and stream lines from airphotos and topographic maps at scales 1:5,000 to 1:50,000 Surface-specific heights and stream lines from ground survey by GPS Remotely sensed heights from airborne and spaceborne photogrammetry, radar and laser altimetry	Detailed terrain parameterization and visualization Estimates of flood inundation, soil moisture and other distributed-parameter hydrological modelling Spatial analysis of terrain and soil properties Terrain-aspect corrections to remotely sensed data Effects of terrain aspect on patterns of solar radiation, evaporation and vegetation
50–200 m	Contours and stream lines from airphotos and topographic maps at scales 1:50,000 to 1:200,000 Surface-specific heights and stream lines digitized from topographic maps at 1:100,000 scale	Broader scale distributed hydrological modelling Geomorphometric regionalization and analysis Sub-catchment analysis for lumped-parameter modelling and assessment of biodiversity
0.2–5 km	Surface-specific heights and stream lines digitized from topographic maps at scales 1:100,000 to 1:250,000 N.B.: Coarse-scale DEMs often are compiled by local averaging of fine-scale data	Height-dependent representations of surface temperature and precipitation Effects of terrain aspect on precipitation and surface roughness on wind Mapping continental drainage divisions
5–500 km	Surface-specific heights digitized from topographic maps at scales 1:250,000 to 1:1,000,000 (see also, N.B., above) National archives of trigonometric points, bench marks and other ground-surveyed terrain heights	Modelling relation of erosion and sediment distribution to tectonism Orographic barriers for general circulation models Broad-scale height and shaded-relief base maps for non-topographic information

Source: Modified after Hutchinson and Gallant (2000)

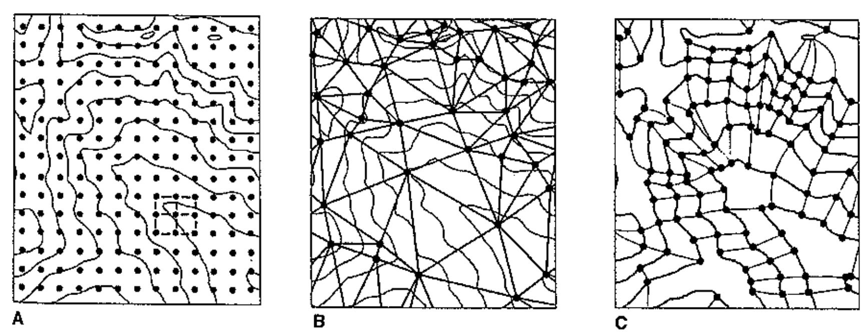

Figure 45 Three contrasting DEM structures for part of a small watershed in California. Dots are height locations. A: square grid, showing 3 × 3 subgrid used for neighbourhood operations; B: triangulated irregular network, TIN; C: intersections of 20-m contours (heavy lines) with slope lines (light). Each panel is 780 m across

are spaced at 10 km. The US National Elevation Dataset (NED) is a seamless 1″ (30 m) DEM (Alaska at 2″) assembled from all 55,000 1:24,000- and 1:63,360-scale topographical maps. Japan and the UK are gridded at 50 m, Italy nominally at 230 m, Australia 9″ (250 m), and Germany and other countries at various resolutions. Distribution of most military DEMs, e.g. DTED for the USA is restricted.

Bypassing contour maps, remote sensing (see REMOTE SENSING IN GEOMORPHOLOGY) can generate DEMs from direct measurements of terrain height (Maune 2001). Technologies include digital photogrammetry, the Global Positioning System (GPS), laser-ranging altimetry (LiDAR), synthetic-aperture radar interferometry (InSAR or IfSAR), thermal emission and reflection radiometry (ASTER), and, for bathymetry, deep-towed SONAR. The 3″ (90 m) DEM compiled from the 2000 Shuttle Radar Topography Mission (SRTM) includes 80 per cent of Earth's land surface (www.jpl.nasa.gov/srtm/). A variably spaced (1–12 km) depth grid, devised from radar altimetry inverted to sea-surface gravity and thence to bathymetry, covers Earth's entire seafloor.

Random and systematic flaws degrade DEM quality, defined by horizontal and vertical accuracy and precision of the constituent heights. Much of the error in DEMs derived from contours originates in the maps themselves, which never were intended to supply data of the high quality desirable for geomorphometry. Because contour maps only approximate terrain, just as DEMs approximate the maps, declared levels of DEM quality are merely statistical; locally, accuracy can be low. Heights expressed in integers can have insufficient vertical precision, or resolution, especially in level terrain. Contour-to-grid processing, always a compromise, is a second source of error. Some interpolation algorithms overrepresent contours; others add spurious terracing, closed depressions, and linear artefacts. Nor do advanced methods assure DEM quality. InSAR, LiDAR, and other remotely sensed data all contain errors, some of them severe, that are unique to their technologies. Where 1″ SRTM data reflect the dense tree canopy, they reproduce the ground surface no more faithfully than the 1″ NED.

Most DEMs must be refined for subsequent analysis (Figure 46). Computer processing can create a TIN or grid DEM from scattered heights

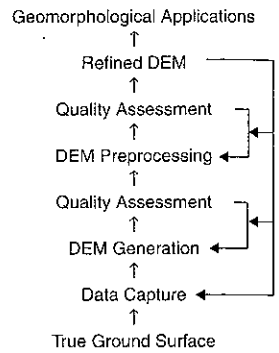

Figure 46 Sequence of activities in preparing a DEM to address geomorphological objectives; modified after Hutchinson and Gallant (2000)

and convert from one data structure or map projection to another. A DEM may be interpolated to finer – within limits, to avoid creating artefacts – or coarser resolutions for compatibility in merging DEMs and registering terrain height to other data. Bulk processing or point-by-point editing corrects errors or replaces parts of a DEM with data of higher quality. Removing long- or short-wave variation from a DEM by digital filtering can enhance detail or subdue erroneous artefacts of production. In creating a grid DEM intended for hydrological application, a drainage-enforcement algorithm reduces digitizing errors and preserves continuity of streams by incorporating channels, ridge lines, and other slope discontinuities (Hutchinson and Gallant 2000).

Suitably preprocessed, grid DEMs are used to describe continuous topography through a spatial calculation adopted from digital image-processing, the neighbourhood operation, wherein a result – for example, relief shading – is obtained from adjacent input values. The input from a grid DEM is a compact matrix of heights, usually 3×3, moved through the data in regular increments; calculations for TINs or contour DEMs are on triangles or quadrilateral facets. Neighbourhood operations characterize terrain in three overlapping domains – RELIEF (Z), spatial (X,Y) and three-dimensional (X,Y,Z).

Most MORPHOMETRIC PROPERTIES derived from DEMs are moment statistics of height Z and its first two derivatives, slope gradient and profile curvature. Spatial parameters of terrain pattern and texture, unreferenced to an absolute datum, are more abstract; common X,Y measures are aspect, the compass direction faced by a slope, and contour curvature. Processing DEMs in the X,Y,Z domain captures the most complex properties – roughness, intervisibility, and variance of relief with azimuth.

Calculations on DEMs both visualize and parameterize the ground surface (Table 11). Digital maps in colour or monochrome portray topography, often in oblique perspective, by shaded relief, slope gradient, or aspect. Multispectral data, symbols for GEOMORPHOLOGICAL MAPPING, and other types of information are commonly displayed as overlays on a base of contoured height.

Among important parameters are the eight DEM derivatives calculated across each continent from the GTOPO30 data (Verdin and Greenlee 1998): a hydrologically integrated DEM, slope gradient and aspect, streamflow direction and accumulation, the topographical wetness index, stream networks, and drainage basins. DEM parameters are used to quantify hillside form (see HILLSLOPE, FORM), map landslide susceptibility, conduct TERRAIN EVALUATION, devise LAND SYSTEMS, assess cross-country trafficability, plan military operations, describe remote submarine and extraterrestrial surfaces, model slope evolution, simulate WATERSHED hydrographs, forecast the extent of flooding, and estimate sediment delivery.

References

Gesch, D.B., Verdin, K.L. and Greenlee, S.K. (1999) New land surface digital elevation model covers the Earth, *Eos, Transactions, American Geophysical Union* 80, 69–70; http://edcdaac.usgs.gov/gtopo30/gtopo30.html

Hutchinson, M.F. and Gallant, J.C. (2000) Digital elevation models and representation of terrain shape, in J.P. Wilson and J.C. Gallant (eds) *Terrain Analysis: Principles and Applications*, 29–50, Chichester: Wiley.

Maune, D. (ed.) (2001) *Digital Elevation Model Technologies and Applications, the DEM Users Manual*, American Society for Photogrammetry and Remote Sensing: Bethesda.

Moore, I.D. (ed.) (1991) Digital terrain modelling in hydrology, *Hydrological Processes* 5.

Verdin, K.L. and Greenlee, S.K. (1998) *HYDRO1k Elevation Derivative Database*, http://edcdaac.usgs.gov/ gtopo30/hydro/index.html

Further reading

Discoe, B. (2002) *The Virtual Terrain Project*, http://www.vterrain.org

Gesch, D.B., Oimoen, M.J., Greenlee, S.K., Nelson, Charles, Steuck, Michael and Tyler, Dean (2002) The National Elevation Dataset, *Photogrammetric Engineering and Remote Sensing* 68, 5–11.

Guth, P.L. (2002) *MicroDEM+ for Windows 95 and NT*, http://www.nadn.navy.mil/Users/oceano /pguth/website/microdem.htm

Kumler, M.P. (1994) An intensive comparison of triangulated irregular networks (TINs) and digital elevation models (DEMs), *Cartographica* 31, 2 (Monograph 45).

Pike, R.J. and Dikau, R. (eds) (1995) Advances in geomorphometry, *Zeitschrift für Geomorphologie, Supplementband* 101.

SEE ALSO: allometry; applied geomorphology; complexity in geomorphology; cross profile, valley; engineering geomorphology; equilibrium slope

RICHARD J. PIKE

DILUVIALISM

A form of CATASTROPHISM in which it is believed that the landscape was shaped by Noah's Flood, as reported in the book of Genesis. Before the true origin of glacial deposits was recognized, such materials, called 'drift' were ascribed by workers such as Buckland (Davies 1969) to a great deluge, when 'waves of translation' covered the Earth. By the 1830s the recognition of the complex stratigraphy of the drift and the discovery of the importance of the Ice Age greatly weakened the diluvial viewpoint.

The term diluvial is still sometimes used in the context, for example, of supposedly water-lain loess deposits.

Reference

Davies, G.L. (1969) *The Earth in Decay*, London: Macdonald.

A.S. GOUDIE

DISSOLUTION

In the geomorphological context, dissolution is the process whereby a rock, or parts of a rock, combine with water to form a solution. As the rock dissolves the different minerals disintegrate into individual ions or molecules and these diffuse into the solution. Hence, study of dissolution must focus on specific minerals as opposed to the rocks that are made up from them. Dissolution of a mineral is congruent when all components dissolve together (i.e. no solid remains) and incongruent where only a part of the components dissolve (for example the alumino-silicate minerals where ions are released in reaction with water but retain most of their elements in re-ordered solids such as kaolinite). There is a very wide range of mineral solubility in water, from gibbsite which is virtually insoluble (0.001 mg l^{-1} at pH 7) through to halite (360 000 mg l^{-1} at pH 7). Rocks made up of minerals with a very low solubility are highly resistant to CHEMICAL WEATHERING, while rocks containing highly soluble minerals such as rock salt are only found at outcrop in the driest places. Between these two extremes are a group of rocks in which dissolution along groundwater flow paths leads to the development of concentrated underground drainage and a landform assemblage known as KARST. Karst develops on silicate and evaporite rocks but is most common on the carbonate rocks, limestone and dolomite. Hence the remainder of this entry discusses the carbonate dissolution process.

The solution chemistry of carbonates is relatively simple as only two major minerals, calcite ($CaCO_3$) and dolomite ($CaMg(CO_3)_2$), are involved. Both are only slightly soluble in pure water (*c.*14 mg l^{-1}) and the solvent action of natural waters depends on their acid content. Organic and mineral acids may be important in some localities, particularly during the earliest (inception) phase of karstification (Lowe *et al.* 2000) but dissolution of calcite and dolomite is generally dominated by carbonic acid produced by hydration of dissolved carbon dioxide. It is frequently stated that the reaction between carbonic acid and 'insoluble' limestone produces calcium bicarbonate which is soluble. However, this is incorrect as there is no evidence for the existence of calcium bicarbonate molecules in solution. In fact there are three elementary chemical reactions in the dissolution of calcite which proceed in parallel:

$$CaCO_3 + H^+ \leftrightarrow Ca^{2+} + HCO_3^- \quad (1)$$

$$2CaCO_3 + H_2CO_3 \leftrightarrow 2Ca^{2+} + 2HCO_3^- \quad (2)$$

$$CaCO_3 + H_2O \leftrightarrow Ca^{2+} + HCO_3^- + OH^- \quad (3)$$

These can be summarized into:

$$CaCO_3 + H_2O + CO_2 \leftrightarrow Ca^{2+} + 2HCO_3^- \quad (4)$$

Similar processes take place in the dissolution of dolomite and are summarized as:

$$CaMg(CO_3)_2 + 2H_2O + 2CO_2 \leftrightarrow Ca^{2+} + Mg^{2+} + Mg^{2+} + 4HCO_3^- \quad (5)$$

The reactions continue until the forward and reverse rates become equal at which point the system is in equilibrium and the solution is said to be saturated with calcite. Addition of any acid to the system will increase the concentration of hydrogen ions and displace the equilibria in a forward direction. This reduces the concentration of CO_3^{2-} and permits more $CaCO_3$ to dissolve so that when equilibrium is re-established the saturated solution has a higher calcium concentration.

In contrast to mechanical erosion processes, these reactions may occur in static water as well as through the range of water velocities. The speed of the reactions, and the amount of mineral dissolved, are controlled by the detailed solution kinetics (discussion of which is beyond the scope of this entry). However, the role of four important factors: carbon dioxide concentrations, temperature, equilibrium conditions and mixing corrosion will be considered briefly.

1. *Carbon dioxide concentrations* The atmospheric concentration of carbon dioxide is close to 0.035 per cent which would yield a saturation value of 70 mg l^{-1} at 10 °C under open system conditions. Observed concentrations are frequently higher and it is generally assumed that this is due to the biogenic carbon dioxide in the soil atmosphere. However, the fluctuations in soil carbon dioxide concentrations are frequently more pronounced than those of calcium concentrations at springs and it is possible that ground air carbon dioxide in the subcutaneous zone may provide a relatively stable source.
2. *Temperature* For any fixed carbon dioxide concentration in a gas mixture in contact with water and rock the calcite solubility decreases with increasing temperature at a rate of approximately 1.3 per cent per degree Celsius. However, this effect is usually less significant than carbon dioxide concentrations in the gas phase and reaction rates, both of which broadly increase with temperature. In addition, regional runoff variations account for a greater proportion of the observed variability in solutional erosion rates than do solute concentration variations.
3. *Equilibrium conditions* The two principal equilibrium conditions under which limestone may be dissolved are the 'open' system in which gas, water and rock are all in contact together such that carbon dioxide is available to replace that used up in the reaction of limestone and carbonic acid, and the 'closed' system in which gas and water come into equilibrium but the gas supply is cut off before contact with rock. Since there is no replacement of carbon dioxide under closed system conditions, the amount of limestone which can be dissolved is less than under open system conditions.
4. *Mixing corrosion* The mixing of two saturated waters produces an unsaturated (aggressive) solution and the mixing of a saturated and an aggressive solution, or of two aggressive solutions, may result in increased aggressivity. In extreme cases the new solution may be capable of dissolving 20 per cent more calcite but 1–2 per cent is more usual in natural waters. Hence, the mixing effect is generally less effective than 'normal' solution and its importance lies in its ability to operate in conditions under which normal solution is impossible, such as narrow fissures and in the phreatic zone.

While the key role of carbonic acid in the dissolution of carbonates was understood by the end of the eighteenth century it was not until towards the end of the twentieth century that details of the equilibrium chemistry of carbonate waters and their importance for speleogenesis and landscape evolution were elucidated, most notably by Dreybrodt (1988, 2000), Palmer (1991) and White (1984).

References

Dreybrodt, W. (1988) *Processes in Karst Systems: Physics, Chemistry and Geology*, Berlin and New York: Springer.

——(2000) Equilibrium chemistry of karst water in limestone terranes, in A. Klimchouk, D.C. Ford, A.N. Palmer and W. Dreybrodt (eds) *Speleogenesis: Evolution of Karst Aquifers*, 126–135, Huntsville, AL: National Speleological Society.

Lowe, D.J., Bottrell, S.J. and Gunn, J. (2000) Some case studies of speleogenesis by sulphuric acid, in A. Klimchouk, D.C. Ford, A.N. Palmer and W. Dreybrodt (eds) *Speleogenesis: Evolution of Karst Aquifers*, 304–308, Huntsville, AL: National Speleological Society.

Palmer, A.N. (1991) The origin and morphology of limestone caves, *Geological Society of America Bulletin* 103, 1–21.

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.

SEE ALSO: corrosion

JOHN GUNN

DIVERGENT EROSION

The term was derived from the evolution of INSELBERGS. When a rock outcrop is laid open in the tropics the rainwater runs off very fast. Thus weathering lacks the moisture for further decomposition. In the surrounding zone the water can percolate into the soil and guarantees continued weathering of the rock there. During the contemporaneous lowering of the surface, the rock outcrop is resistant. Often its sides are further exposed, whilst the neighbouring areas are eroded. Mostly this occurs at a similar rate to that at which the rock outcrop at the base of the regolith is decomposed. Thus the rock grows out relatively slowly from the weathering mantle. Therefore diverging erosion follows diverging weathering. Inselbergs rise up to 300 m and even more. They are developed, however, in all sizes, so that sequences are observed in the tropics. The above derivation developed from these observations. Isolated mountains outside of the tropics are as a rule palaeoforms, sometimes exhumed.

The initial stage of the exposure of the rock outcrops has different causes: thinning of the soil cover occurs in special geomorphic positions, quite often in an area where an ESCARPMENT is originating, i.e. where planational processes produced a slight increase in gradient. The first rock outcrops then evolve into inselbergs which lie in front of or on top of escarpments. Eventually rock outcrops are so numerous that an escarpment develops. Sometimes inselbergs occur near rivers or sea coasts. Rapid downwearing produces an initial small rock outcrop by chance, e.g. by tree fall. Other positions for inselbergs are watersheds, large and small. Examples are the prominent inselbergs in the south of Central Australia: Ayers Rock, the Olgas, and Mt. Connor. All these are developed in sedimentary rocks, which shows that divergent weathering is independent from rock hardness as the lithology is more or less the same in these rocks and the neighbouring plain, at least in areas larger than the inselbergs. Quite often a different spacing of fractures is postulated as a reason for special resistance, but tectonic lines should show repetitive patterns and thus a regular spacing of inselbergs, which is not the case. Divergent erosion is sometimes used for different processes controlled by rock hardness, too, especially in the case of inselbergs. But it is always hard to prove this as rock samples for comparison from the deeply weathered plain are difficult to retrieve.

Rock outcrops are generally resistant in the tropics due to the minor importance of physical weathering and the overwhelming power of chemical processes. The first needs water, if at all, only for a short time, while the second can only work with long wetting, preferably with water containing organic acids. Both are missing on bare slopes. Therefore inselbergs and escarpments in the tropics are often very old. After exposure these rock outcrops are only slightly weathered. Even if special forms like runnels, exfoliation sheets, or small caverns developed, the overall form is not changed. At Ayers Rock these weathering forms are nested. Thus they are of different ages, which proves the stability of the slope during a very long time even under changing climates.

Very steep slopes in Sri Lanka are surprisingly stable. This is not only due to rock outcrops but to the rapid movement of the subterraneous water. For this, soil analysis gives an explanation: in thin sections a very high volume of large pores is seen, as is the relatively stable soil texture due to iron and silica minerals in the matrix or even as cutans on pore walls. Thus the internal water movement has good pathways. Soil stability is maintained due to the low swelling capacity of the kaolinite minerals. The rapid water movement is similar to that on rock outcrops. Thus these processes were called internal divergence. Once the soil fabric is disturbed, e.g. by building a street on a steep slope, water movement is blocked and severe slides may occur.

Divergent erosion as a principle in tropical geomorphology shows a discontinuity of erosion in space and time. It is nearly independent of the forces of gravity but dependent on differences in friction. One can consider divergent erosion as a positive feedback mechanism. A threshold of resistance to weathering and erosion is not surpassed in the case of the exposed rock facets. Thus they possess an extremely low sensitivity to change. The ergodic principle is applicable in the humid tropics where planational lowering is still active to different heights on the periphery of inselbergs due to divergent erosion.

References

Bremer, H. (1972) Flußarbeit, Flächen- und Stufenbildung in den feuchten Tropen, *Zeitschrift für Geomorphologie Supplementband* 14, 21–38.

Bremer, H. and Sander, H. (2002) Inselbergs: geomorphology and geoecology, in S. Porembski and W. Barthlott (eds) *Inselbergs, Biodiversity of Isolated Rock Outcrops in Tropical and Temperate Regions*, Ecological Studies 146, 7–35, Berlin, Heidelberg: Springer.

H. BREMER

DOLINE

Dolines are natural enclosed depressions found in karst landscapes (Ford and Williams 1989). They are subcircular in plan, tens to hundreds of metres in diameter, and can range from a few metres to about a kilometre in width. They are typically a few metres to tens of metres in depth, but some are hundreds of metres deep. Their sides range from gently sloping to vertical, and their overall form can vary from saucer shaped to conical or even cylindrical. Dolines are especially common in terrains underlain by carbonate rocks, and are widespread on evaporites. Some are also found in siliceous rocks such as quartzite. Dolines have long been considered a diagnostic landform of KARST, but this is only partly true. Where there are dolines there is certainly karst, but karst can also be developed subsurface in the hydrogeological network even when no dolines are found on the surface. Dolines have a similar function in karst landscapes to the drainage basin in non-karstic lithologies, in that they drain rainwater from the surface, but in the case of the doline it is discharged underground via an outlet at the lowest point in the doline basin.

The term sinkhole is sometimes used (especially in North America) to refer both to dolines and to depressions where streams sink underground, which in Europe are described by separate terms (including ponor, swallow hole, and stream-sink). Thus the terms doline and sinkhole are not strictly synonymous. Table 12 lists the terms employed by different authors and Figure 47 illustrates six main doline types (both are from and are discussed in more detail in Williams 2003).

Enclosed depressions in karst can be formed by four main mechanisms: DISSOLUTION, collapse, SUFFOSION and regional SUBSIDENCE. In practice the complexity of natural processes often results in more than one mechanism being involved, in which case the doline is polygenetic in origin. A typical case is a depression formed initially by dissolution that later in its development is subject to collapse of its floor into an underlying cave. In such a case, the gentler upper slopes of the doline were formed by dissolution and the steeper lower slopes by collapse.

Solution dolines

The bowl-shaped form of a typical doline indicates that more material has been removed from its centre than from around its margins. Where the principal process responsible for this is dissolution of the bedrock, it follows that there is a mechanism that focuses chemical attack. The amount of limestone that can be removed in solution depends upon two variables: first, the concentration of the solute and, second, the volume of the solvent (in this case the amount of water draining through the doline). Variations in either or both of these variables could be responsible for the focusing of dissolution near the centre of the depression, but if local variation in solute concentration alone were sufficient to explain the occurrence of solution dolines, then they would be found on every type of limestone in a given climatic zone. This is not the case, as illustrated by comparison of landscapes formed on Devonian, Carboniferous, Jurassic and Cretaceous limestones in England, where dolines are most frequently found on Carboniferous limestones and tend to be less prevalent on Cretaceous and Jurassic limestones. It follows, therefore, that local spatial variations in water flow must be responsible for focusing corrosional attack.

The development of dolines of all kinds depends on the ability of water to sink into and flow through karst rocks to outlet springs. The exposure of limestones by erosion provides an input boundary for infiltration of water and a valley incised into the limestone provides an output boundary. Infiltrating rainwater is acidified in the atmosphere and further acidified in the soil. On percolating downwards this water accomplishes most of its dissolutional work within 10 m of the surface. Joints (see JOINTING), faults and bedding-planes vary spatially within the rock because of tectonic history and variations in lithology. Consequently the frequency and interconnectedness of fissures available to transmit flow also varies. Some fissures are more favourable for percolation than others, for example where several joints intersect, and as a result these develop as principal drainage paths. Water flows towards them and as a result they are subjected to still more dissolution by a positive feedback mechanism and so vertical permeability is enhanced. The local surface of water saturation in the EPIKARST is drawn down over the preferred leakage paths similar to cones of depression in the water table over pumped wells; streamlines adjust and resulting flow lines are centripetal and convergent on the preferred drainage zones. By this means solvent flow is focused and, as the surface lowers, the more intensely corroded zones begin to obtain topographic expression as solution dolines. Particularly large solution depressions often occur in the humid tropics where corrosion processes were uninterrupted by Pleistocene glaciations. In these places the term cockpit is sometimes applied to them after a particular style of landscape in Jamaica, where depressions are incised between intervening conical hills.

Although small solution dolines have formed in 15,000 years or so in some mid to high latitude areas that were glaciated in the late Pleistocene, several tens to hundreds of thousands of years are required to develop large solution dolines in limestone. Once formed they may persist in the landscape for several million years provided there is sufficient thickness of limestone for their continued incision. Individual dolines may merge to form compound closed depressions (known as uvalas) and large dolines may subdivide internally into smaller second generation basins. Where all the available space is occupied by depressions, rather like an egg box, the landscape is termed polygonal karst, because the

Table 12 Doline/sinkhole English language nomenclature as used by various authors

Doline-forming processes		Ford and Williams 1989	White 1988	Jennings 1985	Bogli 1980	Sweeting 1972	Culshaw and Waltham 1987	Beck and Sinclair 1986	Other terms in use
Dissolution		solution	solution	solution	solution	solution	solution	solution	
Collapse		collapse	collapse	collapse	collapse (fast) or subsidence (slow)	collapse	collapse	collapse	
Caprock collapse			—	subjacent collapse		solution subsidence	—		interstratal collapse
Dropout	subsidence		cover collapse	subsidence	alluvial	alluvial	subsidence	cover collapse	
Suffosion		suffosion	cover subsidence				—	cover subsidence	ravelled, shakehole
Burial		—	—	—	—	—		—	filled, palaeo-

Source: from Williams 2003 modified from Waltham and Fookes 2003

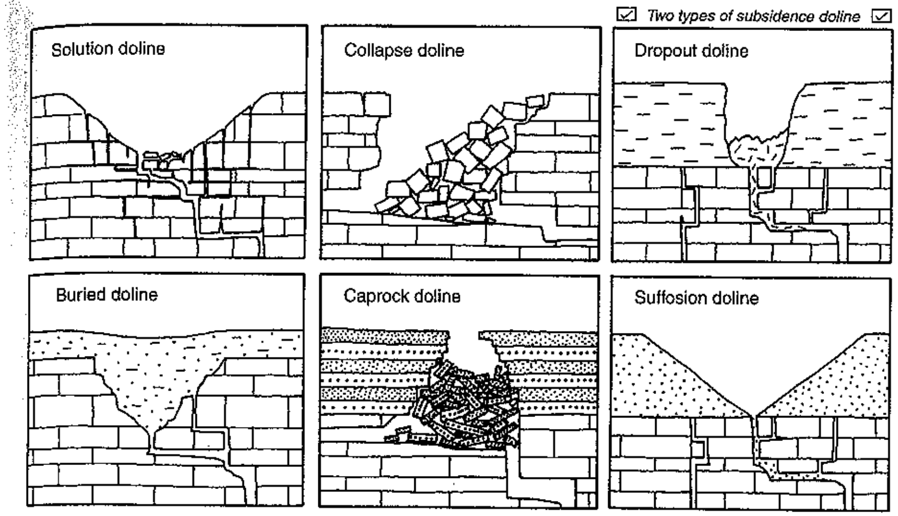

Figure 47 Classification of dolines (after Williams 2003)

topographic divides of the adjoining solution depressions have a polygonal pattern when viewed in plan.

Collapse dolines

Collapse dolines are formed mainly by mechanical processes. There is considerable variation in nomenclature concerning depressions formed mainly by mechanical processes (Table 12), largely because of the variety of materials and processes involved (Waltham 1989). Collapse refers to rapid downward movement of the ground, whereas subsidence refers to gradual movement sometimes without even ripping the surface. These processes can occur in karst bedrock, in caprock that may stratigraphically overly it, and in veneers of unconsolidated sediments. In all cases the collapse has to be preceded by dissolution of the karst rock to form a void into which material can fall. The kind of landforms produced depends upon which of the various materials and processes were involved.

Where collapse dolines form in karst bedrock then the void is commonly part of a cave system. Collapse may occur following undermining from below as the roof of a cavity stopes upwards, ultimately causing the surface above to collapse, or following dissolution from above that weakens the span of a cave roof, causing it to collapse. For example, solutional attack by drainage water near the bottom of a solution doline may combine with upwards stoping of an underlying cave roof to weaken a span from above and below, thereby causing the doline floor to collapse into a cave. Collapse dolines are on average smaller in diameter than solution dolines, although particularly large examples 700 m along their largest axis and up to 400 m deep are known in the Nakanai Mountains of New Britain, Papua New Guinea.

Sometimes a collapse extends from a cave below the modern water-table level, in which case the collapse doline will contain a lake. Such features are known as cenotes after the type-site in the Yucatan Peninsula of Mexico, although similar landforms are found elsewhere, such as in south-east Australia. The deepest known case of a collapse doline containing a lake is the Crveno Jezero (Red Lake) in Croatia, which is 528 m deep from its lowest rim, the bottom of the collapse extending 281 m below the modern level of the nearby Adriatic Sea.

Another process that increases the effective stress on rock arches and subsurface domes is removal of buoyant support by water-table lowering. This increases the effective weight on the span of the roof, resulting in its strength being exceeded and so in its failure and collapse. This occurs because in a fully saturated medium the buoyant force of water is $1\,t\,m^{-3}$, and if the water table is lowered by 30 m, the increase in the effective stress on the rocks is $30\,t\,m^{-3}$. A gradual lowering of the water table occurs with valley incision, because springs are lowered too, and with them the level of the saturated zone that feeds them. More rapid still is the lowering caused by sea-level fall, a process that occurred frequently in the Pleistocene because of repeated glacio-eustatic (vertical movement of sea level caused by glaciation and deglaciation) fluctuations. This particularly affected karsts well connected to the coast such as in Florida, southeastern Australia and Yucatan, where it probably was a significant influence in the development of cenotes.

If unconsolidated coverbeds are drained by water-table lowering, then consolidation and compression occurs, leading to subsiding of the surface and collapse where clastic sediments span de-watered unsupported arches. This is a common process in Florida where porous sandy formations overlie karstified limestones, and has been exacerbated by groundwater pumping for water supplies, which has still further reduced buoyant support. This process and the resulting incidence of collapse attains dangerous hazardous proportions in karstified areas extensively de-watered by mining activities (Beck and Pearson 1995). These dolines in unconsolidated coverbeds are sometimes referred to as cover collapse sinkholes (Table 12).

Subsidence (suffosion/dropout) dolines

When unconsolidated deposits such as alluvium, glacial moraine, loess or sand mantle karstified rock, the sediments are sometimes evacuated downwards through corrosionally enlarged pipes in the underlying karst, resulting in gradual or rapid SUBSIDENCE of the surface. Hence, the term subsidence doline is sometimes used for any closed depression in unconsolidated deposits, although the term is also used for depressions formed by much larger scale regional subsidence. Often a combination of processes is involved in the development of subsidence dolines including corrosion and collapse of the underlying bedrock, as well as suffosion, mudflow and void collapse in the mantling materials. However, the main process by which the sediment moves is known as suffosion and involves the gradual downwashing of fines by a combination of physical and chemical processes. The topographic consequence of this activity depends on whether the material is cohesive or non-cohesive. In cohesive sediments evacuation of material may proceed for some time without any surface expression. However, a void is formed that enlarges and stopes upwards resulting in a sudden, and sometimes catastrophic, failure of the ground surface. The depression thus formed is called a dropout doline or cover collapse doline. In Britain suffosion dolines formed in glacial boulder clay overlying limestone are widely referred to as shakeholes. Similar features but in more uniform finer grained materials are referred to as cover subsidence sinkholes in the USA.

References

Beck, B.F. and Pearson, F.M. (eds) (1995) *Karst Geohazards: Engineering and Environmental Problems in Karst Terrane*, Rotterdam: Balkema.

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

Waltham, A.C. (1989) *Ground Subsidence*, Glasgow: Blackie.

Waltham, A.C. and Fookes, P.G. (2003) Engineering classification of Karst ground conditions, *Quarterly Journal of Engineering Geology and Hydrogeology* 36(2), 101–118.

Williams, P.W. (2003) Dolines, in J. Gunn (ed.) *Encyclopedia of Cave and Karst Science*, London: Routledge.

PAUL W. WILLIAMS

DONGA

Derived from the Nguni word *Udonga*, meaning a wall, it is a term used in southern Africa to describe a gully or BADLAND area caused by severe erosion (Plate 35). Widespread in Lesotho, Zimbabwe, the middleveld of Swaziland, in the Karoo, and Kwazulu-Natal, they are especially prevalent in COLLUVIUM and in deeply weathered bedrock in areas where the mean annual precipitation lies between *c.*600 and 800 mm. Where the materials in which they are developed have high ESP (Exchangeable Sodium Percentage) contents, they may have highly fluted 'organ pipe' sides (Watson *et al.* 1984). Repeated oscillations have taken place in colluvium deposition and palaeosol formation on the one hand, and incision on the other (Botha and Federoff 1995). Causes of incision may include climatic change, and land cover changes brought about by human activities, the latter including the spread of pastoralism and deforestation for iron smelting. Debates about their origin are similar to those that have been raised in connection with the formation of ARROYOS in the American west. Piping (see PIPE AND PIPING) is probably an important process in their development (Rienks *et al.* 2000).

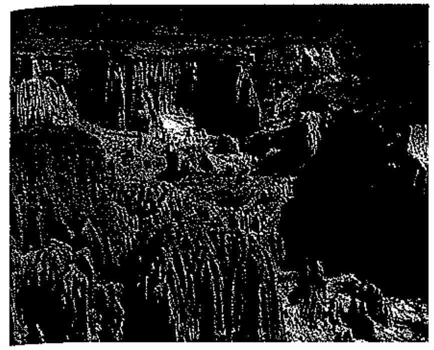

Plate 35 A deep donga developed in highly erodible colluvial material in a valley bottom near St Michael's Mission in central Zimbabwe

References

Botha, G.A. and Federoff, N. (1995) Palaeosols in Late Quaternary colluvium, northern Kwazulu-Natal, South Africa, *Journal of African Earth Sciences* 21, 291–311.

Rienks, S.M., Botha, G.A. and Hughes, J.C. (2000) Some physical and chemical properties of sediments exposed in a gully (donga) in northern Kwazulu-Natal, South Africa, and their relationship to the erodibility of the colluvial layers, *Catena* 39, 11–31.

Watson, A., Goudie, A.S. and Price-Williams, D. (1984) The palaeoenvironmental interpretation of colluvial sediments and palaeosols of the Late Pleistocene Hypothermal in southern Africa, *Palaeogeography, Palaeoclimatology, Palaeoecology* 45, 225–249.

A.S. GOUDIE

DOWNSTREAM FINING

The characteristic decline in the average size of riverbed material with distance downstream is downstream fining. This may include a full sequence from boulder-sized material close to the river source, through gravel-, sand- and silt-sizes to clay-sized material where the river enters the sea. Many rivers do not have all these changes, and may have gravel- or sand-beds at termination. The downstream decline in grain size was first recognized to follow a negative exponential trend by Sternberg (1875) who explained this by the process of ABRASION. Abrasion is significant in many cases, but laboratory measurement of abrasion rates suggests that abrasion alone is insufficient to account for observed rates of fining. It has long been recognized that smaller sediment particles should move more frequently and further than larger ones during BEDLOAD transport. This selective transport mechanism was questioned during the 1980s when it was found that, in GRAVEL-BED RIVERS, there is only slight size selectivity in bedload transport. Further investigation has shown that even a small degree of size selectivity can cause significant downstream fining over long time periods. Downstream fining is thus best explained as a consequence of size sorting during bedload transport, with abrasion and particle breakdown generally acting as secondary effects that may accelerate the rate of fining.

Downstream fining is also one of the downstream adjustments that takes place in graded river systems (see GRADE, CONCEPT OF), along with changes in bed slope, channel width and depth, and flow velocity. The rate of downstream fining (the degree of concavity of a graph of particle size versus distance downstream) is inversely proportional to the length of the river, such that fining is rapid in short rivers and slow in long ones. Close inspection of bed material size data shows that downstream fining is rarely a smooth, continuous process. Abrupt changes in grain size occur where tributaries enter the river, or close to sediment sources (see TERRACE, RIVER). These perturbations are smoothed out at the scale of the whole river, but demonstrate how river networks route both water and sediment downstream, and cause changes in stream ecology. Particularly notable is the abrupt transition from gravel- (>2 mm) to sand-sized (<2 mm) bed material that occurs in many rivers. This transition can result from the supply of large amounts of sand-sized material to the river, or from complex interactions between sediment movement and flow hydraulics that occur as the percentage of sand in the river bed exceeds about 20 per cent.

References

Sternberg, H. (1875) Untersuchungen über Längen-und Querprofil geschiebeführender Flüsse, *Zeitschrift für Bauwesen* 25, 483–506.

Further reading

Hoey, T.B. and Bluck, B.J. (1999) Identifying the controls over downstream fining of river gravels, *Journal of Sedimentary Research* 69, 40–50.

Sambrook Smith, G.H. and Ferguson, R.I. (1995) The gravel-sand transition along river channels, *Journal of Sedimentary Research* A65, 423–430.

SEE ALSO: channels, alluvial; hydraulic geometry

TREVOR B. HOEY

DRAA (MEGADUNE)

Draa, the Arabic word for 'arm', may be used to denote the largest members of the aeolian bedform hierarchy. The term was first used in English by Wilson (1972).

Draas are also known as compound and complex dunes (Breed and Grow 1979), or megadunes (Warren and Allison 1998). They are typically large bedforms with a spacing exceeding 500 m and a height reaching 200 or 300 m and may occur as linear, crescentic, or star forms. Examples of linear draa occur in the Namib Sand Sea, Rub al Khali of Arabia and the Akchar erg of Mauritania; crescentic draa can be found in the Liwa area of the United Arab Emirates and Saudia Arabia, the Namib Sand Sea, and the Algodones dunefield of California. Draa of star form occur in the Grand Erg Occidental and Oriental of northern Africa, the Namib Sand Sea, and the Gran Desierto of Mexico.

Draas are characterized by superimposed bedforms of dune size, with heights up to 10 m and a spacing of up to 300 m. In some places, e.g. the northern Namib Sand Sea, the superimposed dunes appear to be features contemporary with the main bedform (Bristow *et al.* 2000); elsewhere, e.g. in Wahiba Sands of Oman and in Mauritania, the superimposed dunes represent different generations of dunes, in some cases formed in a wind regime different from that which formed the main draa (Warren and Allison 1998; Lancaster *et al.* in press). Thus crescentic dunes may be superimposed on linear draa, and two or more smaller sets of linear dunes are superimposed on older linear draa.

The large size of draa has been thought to be the product of strong winds (e.g. Wilson 1972), but others have suggested that their large size is a product of long continued development in a wind regime that promotes deposition on the dune (e.g. Lancaster 1988). Their large size indicates persistence over long periods of time and reconstitution times in the order of 1 to 100 ka.

Recent stratigraphic and dating studies suggest that some draa (especially linear draa, which tend to conserve their form over long periods) may be composite landforms constructed by multiple generations of aeolian deposition, stability and reworking. In several areas (e.g. UAE, Oman, Mauritania), the cores of large linear draa are at least 15–22 ka old (Glennie and Singhvi 2002; Lancaster *et al.* in press).

References

Breed, C.S. and Grow, T. (1979) Morphology and distribution of dunes in sand seas observed by remote sensing, in E.D. McKee (ed.) *A Study of Global Sand Seas*, United States Geological Survey Professional Paper 1,052, 253–304.
Bristow, C.S., Bailey, S.D. and Lancaster, N. (2000) Sedimentary structure of linear sand dunes, *Nature* 406, 56–59.
Glennie, K.W. and Singhvi, A.K. (2002) Event stratigraphy, palaeoenvironment and chronology of SE Arabian deserts, *Quaternary Science Reviews* 21, 853–869.
Lancaster, N. (1988) Controls of eolian dune size and spacing, *Geology* 16, 972–975.
Lancaster, N., Kocurek, G., Singhvi, A.K., Pandey, V., Deynoux, M., Ghienne, J.-P. and Lo, K. 2003. Late Pleistocene and Holocene dune activity and wind regimes in the western Sahara of Mauritania, *Geology* 30, 991–994.
Warren, A. and Allison, D. (1998) The palaeoenvironmental significance of dune size hierarchies, *Palaeogeography, Palaeoclimatology, Palaeocology* 137, 289–303.
Wilson, I.G. (1972) Aeolian bedforms – their development and origins, *Sedimentology* 19, 173–210.

SEE ALSO: dune, aeolian

NICK LANCASTER

DRAINAGE BASIN

A drainage basin is an area of land that contributes water and sediment to a specific outlet point on a stream. It is separated from other drainage basins by its drainage divide, a boundary that encircles a basin along its highest, outermost ridge tops. The drainage basin is recognized as a fundamental geomorphological unit (Horton 1932: 350) and is frequently used as the primary landscape unit for hydrological, water supply and ecological investigations and for land management activities.

Drainage basins are EROSION created landforms sculpted predominantly by the actions of flowing water. They may be conceptualized as consisting of two geomorphological components: a set of hillslopes dominated by unconfined OVERLAND

FLOW and a branching network of stream channels conveying concentrated flows. The transition from hillslope to channel has been characterized as both indistinct and distinct. Davis (1899: 495) wrote: 'Although the river and hillside waste-sheet do not resemble each other at first sight, they are only the extreme members of a continuous series; and when this generalization is appreciated, one may fairly extend the "river" all over its basin and up to its very divides.' From this perspective, every point within a drainage basin is located along a flow pathway, and a basin is composed of a branching, space-filling drainage network of flow pathways extending from outlet to basin divide. The alternative viewpoint is that the transition from hillslope to channel is determined by a geomorphological threshold (see THRESHOLD, GEOMORPHIC) of channelization that sets a finite scale for dividing a landscape into valleys and hillslopes. Because a drainage basin may be defined upstream of any point on the land's surface, the delineation of a landscape into specific drainage basins is done for some designated purpose.

Several other terms are used synonymously with drainage basin. In Great Britain, catchment is commonly used, whereas in the United States, watershed is a preferred term. Unfortunately, watershed is an ambiguous term that has historically been used as a synonym for drainage divide, and this usage is retained in Great Britain. For large basins drained by a major river, the term river basin is often used (e.g. Amazon River basin). Drainage basin, catchment and watershed do not inherently imply a particular size of drainage area. However, some government agencies in the United States and others are using these terms in size-based classification systems, such as catchment being smaller than a watershed and watershed being smaller than a basin.

The form and structure characteristics of drainage basins and their associated drainage networks are described by their MORPHOMETRIC PROPERTIES, which can be classified into the categories of size, surface, shape, relief and texture. Drainage area, a variable specifying the amount of land area contained within a drainage divide, is an important basin descriptor and is frequently used as a surrogate for the amount of water and sediment yielded by a drainage basin. Because a basin's drainage network is its most prominent feature, network morphometric properties are also used for drainage basin description. A qualitative indication of drainage basin size is indicated by the stream order of its outlet stream (see STREAM ORDERING). The delineation of drainage basins and determination of their morphometric properties has traditionally been done using topographic maps and manual methods. With DIGITAL ELEVATION MODELS (DEM) and geographic information systems (GIS), watershed delineation, drainage network extraction and the automatic calculation of morphometric properties is possible.

Although drainage basins are fundamental geomorphological units, they may not always prove the best choice for organizing landscapes for research or land management purposes. In landscapes dominated by non-fluvial features such as kettle holes (see KETTLE AND KETTLE HOLE) or aeolian dunes (see DUNE, AEOLIAN) drainage networks and drainage basins are often poorly defined and may be difficult to delineate using either manual or GIS methods.

Drainage basin organization

A drainage basin may be organized into two subsidiary landform units: a set of hillslopes and a drainage network. Although hillslopes may occupy 95 per cent or greater of a basin's area, it is the drainage network that noticeably provides the organization of the hillslopes within a basin. The drainage network is the tree-like structure of flow pathways along which water and sediment are concentrated and delivered to the basin outlet. Drainage networks are comprised of exterior and interior links between successive nodes, where nodes are sources, junctions or the basin outlet. Exterior, or first-order links, connect an upstream source node to a downstream junction. Interior links connect two junctions or a junction to the basin outlet.

Using GIS and DEMs, a space-filling network of flow paths can be delineated within a drainage basin, with external links terminating at the basin's exterior divide or internal divides. There are several subsidiary networks contained within the space-filling drainage network. Many investigators discuss the drainage network in terms of streams and stream system, referring to the blue-line streams on topographical maps. Some have considered the stream network to be synonymous with the channel network, but others view channels as geomorphological features identifiable only from field investigation. A drainage basin will also contain a network of VALLEYs, which may or may not contain streams or channels.

Much of the geomorphological analysis devoted to drainage basins has been with respect to the organization and development of the branching link drainage network structure. In a seminal geomorphology paper, Horton (1945) provided many of the concepts supporting modern geomorphological analysis of drainage basins. He provided the basis for the hierarchical method of stream ordering and laws of drainage composition that with later modifications due to Strahler (1957) and others provide a means to organize the understanding about the topologic and geometric properties of drainage networks. In the Horton/Strahler ordering system, source streams (exterior links) are designated as first order. When two first-order streams join, the stream that continues is designated as second order and, in general, at the junction of two streams of equal order, the order of the downstream segment is increased by one. Low-order tributaries may flow into high-order streams without the order being incremented, and the entire section of stream of same order is referred to as one stream segment for the purposes of quantifying the number of streams, stream length, stream slope and contributing area. HORTON'S LAWS of drainage composition refer to the empirical straight-line relationships between these quantities and stream order on semi-log plots.

Horton (1945: 283) also devised the concept of DRAINAGE DENSITY, which indicates the degree of dissection of a drainage basin into subsidiary hillslopes by its channel network. Horton's concepts of network analysis can be applied to any of the drainage networks including channel, valley and GIS-derived networks.

Probabilistic-topologic approaches to network analysis have been devised that examine both the regularity and randomness of drainage networks (Shreve 1966; Smart 1968). The random topology models can readily explain many of Horton's laws. Horton's laws are not actually laws in the strictest sense, but merely expressions of the most probable states of network composition.

Horton's laws characterize the self-similarity in the organization and structure of river networks. This self-similarity has stimulated the use of fractals to characterize river networks (see FRACTAL). Fractals are objects with self-similar geometry, retaining similar organization and complexity over a range of scales. The planform river network when characterized as a fractal has a fractal dimension between one (linear features) and two (filling a two-dimensional space), that can be related to Horton's bifurcation and length ratios (Tarboton *et al.* 1988; La Barbera and Rosso 1989). Hack (1957) first noted an apparent dimensional inconsistency between the lengths of the mainstream and drainage area of a river basin. This can imply elongation with increasing basin size, an idea inconsistent with self-similarity but that has been advanced by some, or that individual streams are themselves fractal with dimension between 1.1 and 1.2. There have been suggestions that space filling is a constraint on the organization of river networks, because in general they should drain an entire two-dimensional area. This leads to a constraint that implies relationships between Horton's length, bifurcation and area ratios and the fractal dimension of individual streams.

The uniting of drainage network configuration and flow characteristics has been proposed in the theoretical framework of the optimal channel network. An optimal channel network is one in which there is energy minimization in the whole and parts of a drainage network. Three principles of optimal channel networks are that energy expenditure in every link is minimized for the transportation of a given discharge, equal energy is expended per unit area everywhere within the network, and energy expenditure is minimized for the network as a whole. A combination of these principles is sufficient to explain the tree-like structure of drainage networks and the empirical relationships describing network organization.

FIRST-ORDER STREAMS and drainage basins are substantial components of river basins. At the upper end of a first-order channel is an unchannelled HILLSLOPE HOLLOW or zero-order drainage basin. Nearly one-half the length of a river basin's drainage network may be contained in its first-order links, and first-order basins can contain 50 per cent of a river basin's area. It is within such small watersheds that runoff produced on hillslopes concentrates into streamflows that initiate channel formation. Low-order basins exhibit the tightest HILLSLOPE-CHANNEL COUPLING and competition between hillslope processes (see HILLSLOPE, PROCESS) and channel processes.

Basin development and evolution

The expression of drainage basin organization is a spatial characterization of drainage basin condition at a point in time. Drainage basins and

networks, however, are not static and change over time due to external influences and internal COMPLEX RESPONSES. To explore drainage basin development, geomorphologists have used three different methods: space-for-time substitution, experimental studies and computer simulation modelling. Early studies of drainage basin evolution were based upon space–time substitution, i.e. the ERGODIC HYPOTHESIS. Maps of different drainage basins in progressive stages of development were ordered to depict a temporally evolving basin undergoing advancing stages of evolution. EXPERIMENTAL GEOMORPHOLOGY has been employed through the use of rainfall simulators raining over 'sandboxes' with drainage system development documented through detailed mapping and time-lapse photography. With advances in computer technology, empirical and theoretical concepts have been implemented into computer models that simulate long-term drainage basin evolution.

Although modes of basin evolution depend upon specific boundary conditions and driving factors, several major steps in basin development can be specified. With the assumption that a basin originates on a smooth surface, its channel network grows through the processes of initiation, elongation and elaboration. After initial development of a skeletal network, a few streams elongate to extend in parallel fashion across much of the length of the surface to form a low drainage density network. Over time, downcutting of these elongated stream channels causes the elaboration of the network through the addition of tributary streams, with the concomitant increase in drainage density. During these initial phases of basin development, drainage density increases rapidly and SEDIMENT LOAD AND YIELD from the basin are high.

Eventually, the channel network reaches a period of maximum extension in which stream elongation through HEADWARD EROSION and infilling by hillslope processes reach an equilibrium condition. Smaller drainage basins become integrated into larger ones through the mechanism of RIVER CAPTURE. As erosion continues and the entire basin drops closer in elevation to BASE LEVEL, the process of network abstraction occurs. Lowered stream slopes reduces STREAM POWER and stream EROSIVITY thereby allowing hillslope processes (see SOIL CREEP and SLOPEWASH) to infill low-order streams and abstract them from the drainage network.

Except in experimental models, rarely will a drainage basin originate on a flat, sloping surface of uniform material, so evolution of real drainage basins will be much more complex than the above model depicts. Also, there is no timescale associated with the evolutionary model described above, but longevity of a drainage basin and its drainage system can be correlated with increasing basin size. Large river basins may persist for tens of millions of years. During such long time frames, changing climatic conditions and tectonic events can so alter conditions that completed evolutionary stages are never fully realized. Channel networks can expand and contract through upstream and downstream migration of channel heads as climate changes over periods of decades to thousands of years. Tectonic events can drop or raise the base level for a drainage basin and initiate a new cycle of headward erosion or halt an existing erosional stage (see TECTONIC GEOMORPHOLOGY).

Also, the channel network evolution model presented above does not account for the causal processes of network development and evolution. Different processes are responsible for channel initiation and evolution in different terrain and environments. In steep terrain, LANDSLIDES may result in the initiation of channels, but in low-gradient basins, headward erosion of the channel head because of changed CONTRIBUTING AREA hydrological conditions may be the primary method of network growth. Though simple descriptive models can specify the overall pattern of evolution, detailed circumstances of basin evolution will be quite variable from one basin to another and may require complex computer simulation models to fully understand.

Geologic and climatic influences

Seeking to explain the regularity exhibited by drainage networks has guided much of the geomorphological interest in these features. Although such explanations provide theoretical bases for network pattern regularity when boundary conditions are uniform in space and invariant over time, actual drainage networks evolve under spatially and temporally varying conditions. Therefore, variability in DRAINAGE PATTERNs can arise because processes defined by geomorphological laws (see LAWS, GEOMORPHOLOGICAL) are operating within non-uniform environmental conditions. In particular, geology and climate have profound effects upon the processes and characteristics of drainage basins and drainage networks. Some have suggested that it would be more beneficial to seek relationships between geology and drainage network characteristics than refine sophisticated theories that disregard such an obvious control (Blöschl and Sivapalan 1995: 282).

The effect of geology upon drainage basin characteristics is difficult to quantify, but geology is nonetheless a controlling factor on drainage basin form and development at multiple scales. Empirical studies have identified relationships between drainage density and bedrock lithology, and rock type can be responsible for many drainage pattern details. Drainage basins underlain by shale or siltstone tend to have higher drainage densities than other lithologies due to low infiltration rates and high production of overland flow. Areas dominated by lithologies with high infiltration rates, such as dune sands, often have poorly defined drainage systems and very low drainage densities. Dendritic stream patterns are common on shales and siltstones, as they are weak rocks that provide limited lithologic resistance to erosion. In geologic formations comprised of stronger rock types, such as sandstones and granites, joint patterns frequently control network pattern because fractured rock along joints has greater ERODIBILITY.

Geologic structures also influence drainage basin shape and drainage network form. At a large scale, river basins may be coincident with geologic or structural basins, with the basin's drainage divide corresponding to the ridgetops of surrounding, uplifted mountain chains. For smaller drainage basins, folded geologic strata can control drainage basin shape and drainage pattern where weaker strata are more readily eroded. Trellis and annular drainage patterns are common in these circumstances. Multiple, low-order streams flow from divides to strike valleys, medium-order streams occupy longitudinal or strike valleys, and master high-order streams run across the strike of more resistant folds in superimposed traversal valleys. Drainage basins may be irregularly shaped with drainage divides following the ridgelines of HOGBACKS or CUESTAS formed by more resistant rock strata. Faults, similar to joints, are areas of weaker, fractured rock and are frequently occupied by streams in superimposed valleys.

Climatic effects upon drainage basin development are promulgated through their controls upon erosion processes. The most critical effect of climate upon drainage basin form and development is through the influence of precipitation and temperature upon vegetation cover. Density of vegetation cover is a predominant control upon erosion and sediment delivery to the drainage network by decreasing soil erodibility. Channel head location, and thereby drainage density, may be dependent upon vegetation because of increased shear stress required for channel formation where vegetation cover is dense. Drainage densities typically are low in arid regions where there is little runoff, are highest in semi-arid regions where sparse vegetation cover does little to prevent channel initiation, are low in moderate precipitation environments where vegetation cover restricts channel development, and can be high even with heavy vegetation cover in areas with high annual rainfall and runoff.

References

Blöschl, G. and Sivapalan, M. (1995) Scale issues in hydrological modeling: a review, *Hydrological Processes* 9, 251–290.

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

Hack, J.T. (1957) *Studies of Longitudinal Stream Profiles in Virginia and Maryland*, Washington: US Geological Survey Professional Paper 294B.

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

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

La Barbera, P. and Rosso, R. (1989) On the fractal dimension of stream networks, *Water Resources Research* 25, 735–741.

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

Smart, J.S. (1968) Statistical properties of stream lengths, *Water Resources Research* 4, 1,001–1,014.

Strahler, A.N. (1957) Quantitative analysis of watershed geomorphology, *Transactions of the American Geophysical Union* 38, 913–920.

Tarboton, D.G., Bras, R.L. and Rodriguez-Iturbe, I. (1988) The fractal nature of river networks, *Water Resources Research* 24, 1,317–1,322.

Further reading

Gregory, K.J. and Walling, D.E. (1973) *Drainage Basin Form and Process*, New York: Wiley.

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

Leopold, L.B., Wolman, M.G. and Miller, J.P. (1964) *Fluvial Processes in Geomorphology*, San Francisco: W.H. Freeman.

Rodiguez-Iturbe, I. and Rinaldo, A. (1997) *Fractal River Basins*, Cambridge: Cambridge University Press.

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

SEE ALSO: GIS

CRAIG N. GOODWIN AND DAVID G. TARBOTON

DRAINAGE DENSITY

Drainage density is defined as the cumulative length of all stream channels in a drainage basin, divided by the drainage basin area. The dimension of drainage density is the inverse of length. In principle, drainage density is a fundamental measure of landscape dissection and half its reciprocal is an average measure of length of overland flow (or distance from divide to the nearest channel). Drainage density has been shown to vary as a function of climate, past climate conditions, biomass, parent material, lithology, relief, time and land use. Unfortunately, no consistent relations have been demonstrated from region to region and Schumm (1997) comes to some discouraging conclusions from a number of detailed basin studies. Nevertheless, drainage density is an important geometric parameter for channel networks as it determines the spacing of channels, the length of hill slopes, the maximum length scale of slope failures and reflects the processes governing landscape dissection. The hydrological response of a channel network is strongly influenced by drainage density and sediment erosion rates have been linked to channel spacing. There are relatively few studies of how drainage density varies with time. Flume table experiments (Schumm *et al.* 1987), studies of land fills (Schumm 1956), glacial tills (Ruhe 1952), coastal terraces (Kashiwaya 1987) and drainage networks on an anticlinal fold which has been progressively uplifted during the past 250,000 years (Talling and Sowter 1999) are representative examples of such studies.

It should be pointed out that the definition of drainage density begs at least two questions:

1 *How is a stream channel defined?* This is a deceptively simple question which does not have a simple answer. Montgomery and Dietrich (1992) suggest that there is an empirically defined topographic threshold for channel head locations which defines the boundary between essentially smooth and undissected slopes and the valley bottoms to which they drain. They derived an experimental relation of drainage area versus local slope for channel heads, unchannelled valleys and low-order channel networks from different study areas. Local slope was measured in the field and drainage area was determined from topographic maps. High slopes generate channels from smaller basin areas and lower slopes require larger basin areas to produce a channel. At the same time, spatial heterogeneity, reflecting the controlling factors listed above, introduces variability into these relations. An empirical, field-based definition of a channel uses the presence of fluvial incision and one or two stream banks, but finger tip tributaries are often indeterminate in the field.

2 *What is the relation between stream channels drawn on maps or stream channels detectable on air photographs and actual stream channels on the ground?* There is a basic stream channel network, which is composed of perennial streams and which expands and contracts with runoff, or there is the active channel network, which is composed of ephemeral, intermittent and perennial streams. In addition, the use of contour crenulations as evidence of the presence of channels will result in the inclusion of parts of valleys that do not contain active channels. Clearly, the largest problems concern the uppermost finger tip tributaries of a drainage basin. When air photos are used, there are further problems of visibility below tree canopies and as always, the scale of the photograph or the map will be a constraint on the resolution achievable. In sum, what is measured by one investigator may not be the same phenomenon as that which is measured by another.

If we assume that the identification and measurement problem can be resolved, a variety of theoretical issues relating to drainage basin characterization and evolution can be broached. Strahler (1956) in developing his view of the drainage basin as an open system tending to achieve a steady state of operation asked how to predict erosional or aggradational response by drainage basins when land use or climate changed. Central to his theoretical discussion was the role of drainage density. He argued that because drainage density is the most valuable scale index with respect to degree of dissection of a basin that it should be possible to express drainage density as a function of several variables that control the evolution of the basin. These variables he deduced in part from Horton (1945) as runoff intensity, an erosion proportionality factor, slope gradient, relief, kinematic viscosity of runoff and acceleration of gravity. Through application of the Buckingham Pi Theorem, he reduced the equation to a function containing four dimensionless groups:

1 the product of drainage density and relief (the ruggedness number)
2 the product of runoff intensity, erosion proportionality factor and slope gradient (the Horton number)
3 the product of runoff intensity, kinematic viscosity and relief (a basin Reynolds number)
4 the square of runoff intensity divided by relief times acceleration of gravity (a basin Froude number).

By solving for drainage density, drainage density is shown to be inversely proportional to relief times a function of the Horton, Reynolds and Froude numbers. The challenge of solving this function has still not been met, though the topic of drainage basin transformation has been put onto a more rational basis.

Melton followed up Strahler's analysis with one paper on drainage basin growth models (1958a) and another on the theory of variable systems (1958b), both of which relied heavily on the assumption of the importance of drainage density. Melton (1958a) demonstrated a close relation between F (stream frequency = number of streams per unit area) and D (drainage density) for mature basins with a wide range of orders, valley side slope angles, climates and rock types. The relation, subsequently known as Melton's Law, is of the form $F = 0.694\ D^2$. Shreve (1967) revisited this relation using links instead of streams and found a related term $K = 0.667$ for topologically random networks. The dimensionless ratio F/D^2 varies inversely with valley side slope and basin relief (where area and channel length are held constant) and is interpreted as a measure of the completeness with which a channel system fills a basin outline. For an ideal basin of $1\ \mathrm{mi}^2$, Melton's Law is postulated to be a growth model. This argument is predicated on the assumption that many basins measured at one point in time can be considered equivalent to the behaviour of a single basin over time. The approach taken in Melton (1958b) is different. He arranges fifteen variables of geomorphic, surficial and climatic elements into two related variable systems on the basis of correlation coefficients for every possible pair in the study of 59 drainage basins. 'Melton's ambitious field program of data collection, coupled with his analysis of the interrelations of the components of a drainage basin and the variables that influence morphology, is a model for future geomorphic studies' (Schumm 1977: 180). The high correlation of drainage density with per cent bare area and a precipitation effectiveness index fits well with the Horton theory of drainage density as a function of the resistivity of the surface to erosive forces, determined in part by vegetation which in turn determines the mean length of overland flow. Keylock (personal communication) has shown that the most frequently cited of Melton's contributions are Melton's Law (1958a) and his correlation structure approach to geomorphology (Melton 1958b), both of which emphasize the importance of drainage density.

References

Horton, R.E. (1945) Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology, *Geological Society of America Bulletin* 50, 275–370.

Kashiwaya, K. (1987) Theoretical investigation of the time variation of drainage density, *Earth Surface Processes and Landforms* 12, 39–46.

Melton, M.A. (1958a) Geometric properties of mature drainage systems and their representation in an E4 phase space, *Journal of Geology* 66, 35–56.

——(1958b) Correlation structure of morphometric properties of drainage systems and their controlling agents, *Journal of Geology* 66, 442–460.

Montgomery, D.R. and Dietrich, W.E. (1992) Channel initiation and the problem of landscape scale, *Science* 255, 826–830.

Ruhe, R.V. (1952) Topographic discontinuities of the Des Moines lobe, *American Journal of Science* 250, 46–56.

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

——(1977) *Drainage Basin Morphology*, Pennsylvania: Dowden, Hutchinson and Ross.

——(1997) Drainage density: problems of prediction and application, in D.R. Stoddart (ed.) *Process and Form in Geomorphology*, 15–45, London: Routledge.

Schumm, S.A., Mosley, M.P. and Weaver, W.E. (1987) *Experimental Fluvial Geomorphology*, New York: Wiley.

Shreve, R.L. (1967) Infinite topologically random channel networks, *Journal of Geology* 75, 178–186.

Strahler, A.N. (1956) The nature of induced erosion and aggradation, in W.L. Thomas (ed.) *Man's Role in Changing the Face of the Earth*, 621–638, Chicago: University of Chicago Press.

Talling, P.J. and Sowter, M.J. (1999) Drainage density on progressively tilted surfaces with different gradients, Wheeler Ridge, California, *Earth Surface Processes and Landforms* 24, 809–824.

SEE ALSO: dynamic geomorphology; stream ordering

OLAV SLAYMAKER

DRAINAGE PATTERN

Because river channels concentrate surface flow and erode into landscape more efficiently than other processes, new channels tend to persist from the pattern initially developed and are subsequently hard to alter. A collection of river channels joined together is called a drainage network, how it is laid out on the ground in plan view is called the drainage pattern, and the channels together with all the land surface that drains to the channel is called the DRAINAGE BASIN. Channels when they join normally do so in an accordant manner, the channels join without a sudden break in elevation (sometimes called Playfair's Law), unless they occupy unmodified glacial terrain, in which case a discordant junction is called a HANGING VALLEY. Subsequent adjustments to networks and patterns may occur when rivers are close together, and the divides between them may be broached by erosion or overflow, or underground drainage may divert water from one system to another prior to there being a surface connection of the rivers. Exploitation of geological weakness by surface erosion eventually causes the overall drainage pattern to reflect the patterns of weakness in the underlying rocks. Major joints and fracture zones may influence subsurface as well as surface drainage and tend to localize major channels. Adjustments by divide erosion and breaching (river piracy, RIVER CAPTURE, diversion) will be most common early in the history of a landscape when relative RELIEF is least. Adjustments by underground diversion may take longer to become active features because large subterranean networks, usually developed in soluble bedrock such as carbonates (KARST terrains), are needed to divert substantial drainage (abstraction). Subterranean diversion is favoured by increasing local relief in the drainage which may permit steeper hydraulic gradients between adjacent channels. Drainage patterns which derive their water entirely from regions external to the locality in question – such as the Nile River in Egypt – are called exoreic, and systems which drain to a central closed depression such as the Jordan River to the Dead Sea, and the basin draining to the Great Salt Lake in Utah – are endoreic.

The nineteenth and early twentieth-century geomorphologist W.M. Davis (1889, 1899) developed an elaborate scheme to describe the components of a river drainage network as they related to stages in its physiographic development. Of those terms, those which remain in common use are *consequent* and *subsequent*. *Consequent* streams are those that develop on the initial land surface in response to regional slope and any random surface declivities. Because they must eventually follow regional slope they usually reflect the tectonic framework of uplift, rather than details of the underlying geology. The term has normally been applied to large streams, but can also describe initial drainage on any new surface – such as recently glaciated terrain. *Subsequent* streams describe streams which, through geologic time, have been able to exploit differences in the relative erodibility of the underlying geology as the drainage system incises slowly into the uplifted block of land. Typically they develop along the geological strike exploiting, for example, weak shales or clays exposed between stronger formations (e.g. sandstones or limestones) in a sequence of sedimentary rocks so that long continued weathering and subaerial erosion over CYCLIC TIME etches out a skeleton of the underlying geology – thus the ridges and valleys of the Appalachian Ranges along the eastern side of North America reveal the folded structures; less dramatically the valleys and escarpments of southern England and northern France also reveal the geological structures. The effect is even more dramatic in dry climates with no masking vegetation. In igneous and metamorphic terrain master joints and shear zones may provide weakness to exploit (Figure 48c). Faults and fault zones, with heavily fractured rocks allowing access for weathering agents, are often weak zones in any geological terrain.

Because consequent drainage flows down the regional slope regardless of local variations in geology, such streams are often used to reconstruct the initial stages of a landscape. However,

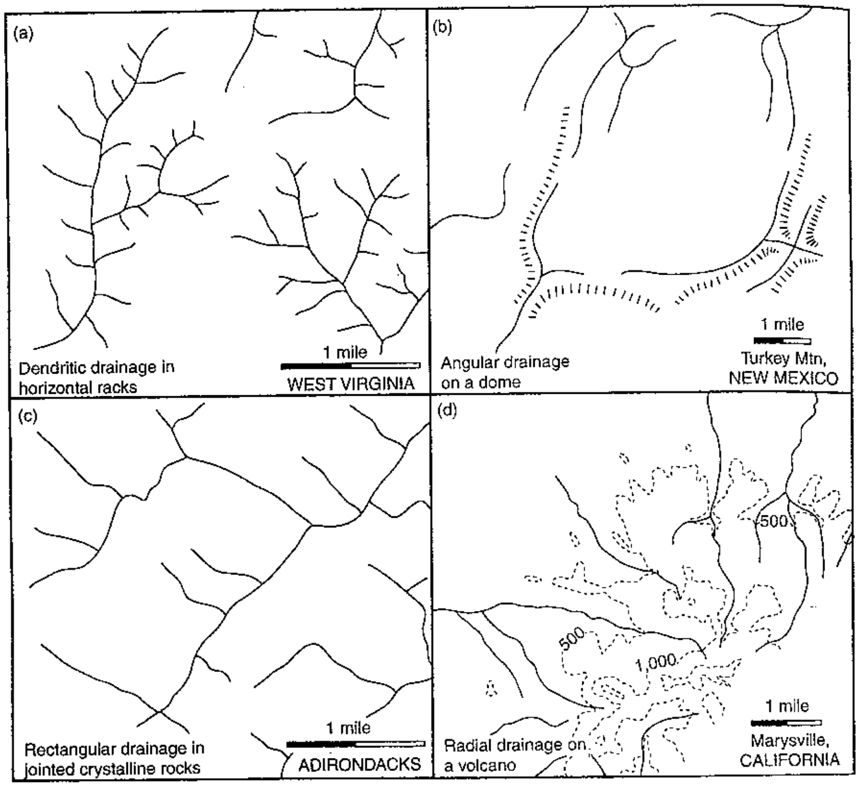

Figure 48 Drainage patterns in relation to topography and geological structures

even consequent streams can be disrupted by continued uplift, with geological structures growing upwards into the overlying streams. If the river channel can erode its bed fast enough to maintain a continuous downslope against the rising land, the river is called antecedent. As a result a river may be seen to have cut a channel, often seen as a deep gorge (see GORGE AND RAVINE), through a prominent topographic ridge around which it might have otherwise been forced to flow. The north to south segments of the Ganges and the Brahmaputra in the Himalaya have been cut in response to, and across the rising folds of, the mountain system.

On occasion though, the rising structure blocks the channel and causes upstream ponding, whose new outlet may provide an entirely new pattern. Complete reversal is possible too. The Amazon originally drained to the Pacific, but its course was reversed by the rising Andean ranges. A related condition, however, is when a regional river system, developed for example on gently tilted sedimentary strata, slowly erodes away that sediment and then erodes into a very different geological underlay. If the sedimentary cover rocks are lying unconformably upon the rocks below, the drainage pattern is said to be superimposed or superposed (Tarr 1890). It is doubtful in practice that either antecedence or superimposition are ever pure conditions because rarely can the full tectonic history of the region be known (Smith *et al.* 1999).

Also, large-scale topographic patterns characterizing the initial topography of the area may be reflected, as for example with radial drainage, such as in the English Lake District where original drainage lines have been greatly accentuated by glacial deepening. Part of a miniature example of radial drainage developed on a volcano is shown in Figure 48d.

Davis developed many terms for other parts of the drainage system as they related to a supposed sequence of drainage and landscape development, and with respect to the original regional slope. These other terms are: insequent, resquent, obsequent; but they have fallen into disuse. Full definitions are available in Lobeck (1939: 171). Of these, insequent streams describe the myriad of streams for which no discernible control can be detected, and which give rise to dendritic patterns (Figure 48a).

Despite the variations in apparent patterns (Figure 48) the patterns that matter most to the operation of the system are the internal structure of connections, and the plan of the DRAINAGE BASIN on the ground. Circular basins concentrate flow more rapidly, and generate larger peak flows than elongate basins. The structural arrangement of channels tends to reflect that of the ground plan – dendritic or vein-like structures being found usually in oval and round basins with homogeneous bedrock (Figure 48a). The Kentucky region with nearly level sedimentary rocks, and lying beyond the glacial limit, has often been used as a basis for comparison with random or randomly generated drainage networks (Mark 1983).

References

Davis, W.M. (1889) The rivers and valleys of Pennsylvania, *National Geographic Magazine* 1, 183–253.
——(1899) The Geographical Cycle, *Geographical Journal* 14, 481–504.
Lobeck, A.K. (1939) *Geomorphology: An Introduction to the Study of Landscapes*, New York: McGraw-Hill.
Mark, D.M. (1983) Relations between field-surveyed channel networks and map-based geomorphometric measures, Inez, Kentucky, *Annals of the Association of American Geographers* 73(3), 358–371.
Smith, B.J., Whalley, W.B. and Warke, P.A. (1999) *Uplift, Erosion and Stability: Perspectives on Long Term Landscape Development*, Geological Society Special Publications, 162, London: Geological Society of London.
Tarr, R.S. (1890) Superimposition of drainage in central Texas, *American Journal of Science* 3rd Series 40, 359–361.

Further reading

Brown, E.H. (1960) *The Relief and Drainage of Wales*, Cardiff: University of Wales.
Johnson, D.W. (1927) *Stream Sculpture on the Atlantic Slope*, New York: Columbia University Press.
Woolridge, S.W. and Linton, D.L. (1955) *Structure, Surface and Drainage in South-East England*, Institute of British Geographers, No. 11, London: George Philip.

KEITH J. TINKLER

DRAPE, SILT AND MUD

A thin deposit of waterlain silt or mud coating a pre-existing morphological feature. Drapes generally grade upwards, sometimes exhibiting internal laminations. They are typically several centimetres thick, though their form and composition vary spatially and temporally, in response to factors including sediment supply, and the hydrological (fluvial or current) regime.

Drapes are indicative of tidal/subtidal settings and are considered one of the most distinctive features of such an environment. Typical tidal regimes exhibit ebb-flood cycles in which one current is more dominant than the other. During periods of tidal dominance various bedforms are produced (e.g. sand bars, ripples and dunes) that are characteristic of the tidal regime. However, there is a period of time during high tide and low tide where no dominant direction of flow exists (termed the slackwater period). During this short period, the suspended sediment of the water may fall and settle on the pre-existing features formed during the dominant tidal period. The subsequent current stage may partly rework the mud or clay drape producing an erosive reactivation surface, though the cohesive nature of the fine clay-rich drapes commonly protects against tidal erosion and preserves the drape. Over time, continued preservation of alternating tidal (sand deposited) and slackwater deposits (mud/clay drapes) produces sand/mud couplets, also called tidal rhythmites. This systematic deposition of tidal rhythmites has allowed detailed reconstruction of past tidal regimes (e.g. Visser 1980), and are particularly distinctive of inshore tidal environments.

Drapes may also form in fluvial environments, particularly within rivers that exhibit seasonal flow and flooding. As flooding wanes, the clay/mud settles on river levees, etc. thus signifying slackwater periods and forming drapes.

Reference

Visser, M.J. (1980) Neap-spring cycles reflected in Holocene subtidal large-scale bedform deposits: a preliminary note, *Geology* 8, 543–546.

STEVE WARD

DRUMLIN

Drumlins are roughly ovoid-shaped hills dominantly composed of glacial debris. Typically, they occur within groups or fields of several thousands, exhibit strong, *en echelon*, long-axis preferred orientation paralleling the main direction of ice flow. The classical shaped drumlin usually has a steeper stoss end and a tapered lee-side, however variants on this shape are perhaps more common than the classical shape itself. Drumlins may vary from 5 to 200 m in height, 10 to 100 m in width, and overall from 100 m to several kilometres in length (see e.g. Mills 1987). Interestingly, few modern drumlins appear from beneath modern-day ice masses other than on James Ross Island, Antarctica, in the proglacial areas of some Icelandic outlet glaciers, and at the Bifertensgletscher, Switzerland. Vast drumlin fields – numbering in thousands – exist, for example, in Canada, Estonia, Finland, Ireland, Germany, Poland, Russia and the USA. The topographic locations within which drumlins are found are many and varied. Drumlins occur in both lowland and highland terrains, beneath ice sheets and valley glaciers, close to terminal MORAINEs and may, in places, appear contiguous with these moraines, while elsewhere they occur on the edge of ice sheet centres. Occasionally, a radiating pattern can be observed within a drumlin field that has been interpreted as evidence of basal crevasse infilling owing to divergent ice flow close to an ice margin. It has also been suggested that drumlins, in association with Rogen and fluted moraines, may be related to deformable beds beneath ice sheets, linked to fast basal ice ($>500\,\mathrm{m\,a^{-1}\,mm}$) and a preferential location within ice streams. Limited relationships appear to occur between drumlins and topography.

Drumlins are composed of a vast range of sediment types of varied provenance, containing an array of sediment structures and forms (Figure 49). In the past, drumlins were mistakenly perceived as being composed almost exclusively of subglacial tills; although many drumlins contain stratified sediment. Drumlins composed of stratified sediments are known, for example, near Velva, North Dakota and Livingstone Lake, Saskatchewan. Also, stratified sediments can occur in individual drumlins that often 'sit' adjacent to till drumlins as in Peterborough, Ontario. Many drumlins have observable cores of bedrock, boulder dykes and other non-glacial nuclei around which subglacial debris has accreted or been emplaced. In some cases, drumlin or drumlinoidal forms can be observed 'carved' from bedrock in the form of roc-drumlins. However, most drumlins do not appear to have obvious cores around which they have been 'built' and these forms remain enigmatic in origin.

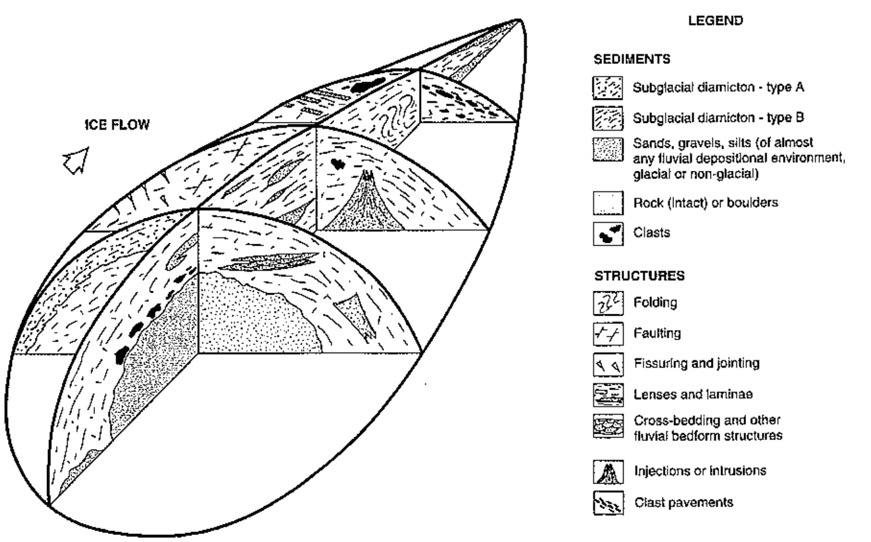

Figure 49 A general model of internal sediments and structures found within drumlins

Clast fabrics within drumlins appear, in some cases, to follow the outer morphology of the drumlin, while others exhibit a transverse, 'herringbone' style pattern. In many cases the complexity of internal sedimentological structures provides a random fabric orientation. Drumlins exhibit such a wide complexity of form and internal composition that it is impossible to characterize an 'ideal' drumlin. Many drumlins are found lying on top or obliquely across other larger drumlin forms (megadrumlins). Drumlin shapes may vary enormously and may reflect formative processes or simply post-depositional subaerial mass movement. Many drumlin fields progressively change as part of a continuum of bedforms, thus drumlin genesis would appear tied to subglacial environments conducive to Rogen and fluted moraine formation. The question of drumlin formation has attracted an array of research. In terms of drumlin formation, it is germane to consider the 'conditions' that must be met by any formative hypotheses, assuming that a single explanation does exist for such a diverse landform/bedform type.

Any explanation of drumlin formation must address the following issues: (1) the diverse location of drumlins and their propensity to occur in 'fields'; (2) the differing shape and morphology of drumlins; (3) the range of sediment types and structures within drumlins; (4) the existence of rock-cored and non-rock-cored drumlins, often in proximity to each other; (5) the presence of drumlins in bedform continua in some, but not all, cases; (6) the relationship of drumlins to subglacial glacidynamics and hydraulics; (7) the chronology of drumlin formation whether drumlins form simultaneously as a single field or develop into a field by repeated 'overprinting' in a single glacial phase or repetition over several glacial phases; (8) stages of drumlin development whether formed *en masse* or by gradual accretion in a single continuous event or interrupted accretionary events; and, finally, (9) a 'trigger' mechanism(s) that is operative in certain specific conditions yet not under others.

At present, three main groups of drumlin-forming hypotheses can be identified:

1. By moulding of previously deposited material within a subglacial environment in which a limited amount of subglacial meltwater activity occurs (possibly where a frozen bed transforms to a melted bed; Menzies 2002). Meltwater may influence moulding and deformational processes by acting either as a lubricating basal film at the upper ice–bed interface, or as porewater reducing subglacial sediment effective stresses. Debris is moulded by direct deformation of previously deposited sediment (both glacial and non-glacial) into drumlinoidal shapes by basal ice contact following smearing-on or sculpting process(es).
2. By anisotropic differences in the subglacial debris under melted-bed conditions owing to: (a) dilatancy (Smalley and Unwin 1968); (b) porewater dissipation; (c) localized freezing; (d) helicoidal basal ice flow patterns (Aario 1977); or (e) localized subglacial debris deformation (Boulton 1987; Menzies 1989). Within this specific group, meltwater activity is of limited impact, whereas porewater is considered critical in local bed debris deformation. Changing stress field and/or stress/strain histories owing to transient basal glacidynamics locally affecting subglacial debris rheology are the important parameters in determining whether drumlins begin to form or not.
3. By the influence of active basal meltwater (under frozen bed conditions) carving cavities beneath an ice mass and later infilling with assorted but predominantly stratified sediment or by the subglacial meltwater erosion of already deposited sediment at the upper ice–bed interface (Dardis and McCabe 1987; Sharpe 1988), or through the entire drumlin, or the sculpting by fluvial processes of previously deposited sediment (Shaw *et al.* 2000). This hypothesis demands meltwater flow of catastrophic discharges from beneath certain areas of an ice mass across the upper ice–bed interface yet permitting the overall ice mass to remain glacidynamically stable. This form of

drumlin development, as with the hypothesis in (1), requires a two-stage process of initiation, beginning with either a pre-formed cavity or pre-existing sediment at the upper ice–bed interface. The latter stage need not be linked directly to the former stage, therefore in some cases although conditions may be suitable for initiation for the first stage, the second stage may not continue toward the critical point (trigger) of drumlin development.

In all these hypotheses the conditions at the subglacial interface(s) are the key to subsequent drumlin formation and, in the long-term, to drumlin 'survival'. A complex relationship must exist between basal glacidynamics, subglacial sediment rheology, and hydraulics for any particular area of ice bed. Fluctuations in state or stress levels or meltwater production and pathways will affect all other parameters to some extent. Certain fluctuations may cross critical thresholds that cannot be reversed, while others may exhibit varying degrees of hysteresis. The likelihood or otherwise of subglacial conditions occurring in any or all these hypotheses remains a fundamental, ongoing, research problem.

References

Aario, R. (1977) Associations of flutings, drumlins, hummocks and transverse ridges, *Geojournal* 1, 65–72.

Boulton, G.S. (1987) A theory of drumlin formation by subglacial sediment deformation, in J. Menzies and J. Rose (eds) *Drumlin Symposium*, 25–80, Rotterdam: A.A. Balkema.

Dardis, G.F. and McCabe, A.M. (1987) Subglacial sheetwash and debris flow deposits in Late Pleistocene drumlins, Northern Ireland, in J. Menzies and J. Rose (eds) *Drumlin Symposium*, 225–240, Rotterdam: A.A. Balkema.

Glückert, G. (1973) Two large drumlin fields in central Finland, *Fennia* 120.

Menzies, J. (1989) Subglacial hydraulic conditions and their possible impact upon subglacial bed formation, *Sedimentary Geology* 62, 125–150.

——(2002) Ice flow and hydrology, in J. Menzies, (ed.) *Modern and Past Glacial Environments*, 79–130, Oxford: Butterworth-Heinemann.

Mills, H.H. (1987) Morphometry of drumlins in the northeastern and north-central USA, in J. Menzies and J. Rose (eds) *Drumlin Symposium*, 131–147, Rotterdam: A.A. Balkema.

Sharpe, D.R. (1988) Late glacial landforms of Wollaston Peninsula, Victoria Island, Northwest Territories: product of ice-marginal retreat, surge, and mass stagnation, *Canadian Journal of Earth Sciences* 25, 262–279.

Shaw, J., Faragini, D.M., Kvill, D.R. and Rains, R.B. (2000) The Saythabasca fluting field, Alberta, Canada: implications for the formation of large-scale fluting (erosional lineations), *Quaternary Science Reviews* 19, 959–980.

Smalley, I.J. and Unwin, D.J. (1968) Formation and shape of drumlins and their orientation and distribution, *Journal of Glaciology* 7, 377–480.

JOHN MENZIES

DRY VALLEY

A valley which is seldom, if ever at the present time, occupied by an active stream channel. Such valleys occur in a wide range of climatic and lithological environments, including extensive areas of Britain and Europe, where they have often been regarded as a product of intense incision under former periglacial conditions (Büdel 1982). There have been many different hypotheses put forward to explain why such valleys are dry (see Table 13; Goudie 1993).

The uniformitarian hypotheses require no major changes of climate or base level, merely the operation of normal processes through time; the marine hypotheses are related to base-level changes; and the palaeoclimatic hypotheses are associated primarily with the major climatic changes of the Pleistocene. British dry valleys (Plate 36) show a considerable range of shapes and sizes, from mere indentations in escarpments, to great winding chasms like Cheddar Gorge in the Mendips. Many, but not all, are formed in carbonate rocks.

Other dry valleys include those that occur in the world's warm deserts and which are relicts of former pluvial conditions and of extensive groundwater sapping (e.g. the MEKGACHA of the Kalahari (Nash 1996). At the other extreme, there are the famous dry valleys of Antarctica, which were cut by former outlet glaciers draining from the Polar Plateau (Summerfield *et al.* 1999).

References

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

Goudie, A.S. (1993) *The Nature of the Environment*, Oxford: Blackwell.

Nash, D.J. (1996) On the dry valleys of the Kalahari: documentary evidence for environmental change in central southern Africa, *Journal of African Earth Sciences* 27, 11–25.

Table 13 Hypotheses of dry valley formation

Uniformitarian	
1	Superimposition from a cover of impermeable rocks or sediments
2	Joint enlargement by solution through time
3	Cutting down of major through-flowing streams
4	Reduction in catchment area and groundwater lowering through scarp retreat
5	Cavern collapse
6	River capture
7	Rare events of extreme magnitude
Marine	
1	Non-adjustment of streams to a falling Pleistocene sea level and associated fall of groundwater levels
2	Tidal scour in association with former estuarine conditions
Palaeoclimatic	
1	Overflow from proglacial lakes
2	Glacial scour
3	Erosion by glacial meltwater
4	Reduced evaporation caused by lower temperatures
5	Spring snowmelt under periglacial conditions
6	Runoff from impermeable permafrost

Plate 36 A dry valley, the Manger, developed in the Vale of the White Horse near Wantage, southern England. The valley is developed in Cretaceous chalk and on the left side shows a series of parallel flutes or dells, which some investigators have proposed to be avalanche chutes

Summerfield, M.A., Stuart, F.M., Cockburn, H.A.P., Sugden, D.E., Denton, G.H., Dunai, T. and Marchant, D.R. (1999) Long-term rates of denudation in the Dry Valleys, Transantarctic Mountains, Southern Victoria Land, Antarctica based on in-situ-produced cosmogenic nuclides, *Geomorphology* 27, 113–129.

A.S. GOUDIE

DUNE, AEOLIAN

Aeolian dunes form part of a hierarchical system of bedforms developed in wind-transported sand which comprises: (1) wind ripples (spacing 0.1–1 m); (2) individual simple dunes or superimposed dunes on draa or compound and complex dunes (spacing 50–500 m); and (3) draa or compound and complex dunes (spacing >500 m). Most dunes occur in contiguous areas of aeolian deposits called ergs or sand seas (with an area of $>100\,km^2$). Smaller areas of dunes are called dunefields (see SAND SEA AND DUNEFIELD). The majority of dunes are composed of quartz and feldspar grains of sand size, although dunes composed of gypsum, carbonate and volcaniclastic sand as well as clay pellets also occur.

The formation of areas of dunes is determined by the production of sediment of a range of suitable particle sizes, the availability of this sediment for transport by wind and the transport capacity of the wind (Kocurek and Lancaster 1999). Most dunes are derived from material that has been transported by fluvial or littoral processes. Important sources include marine and lacustrine beaches, dry lake basins, river floodplains and deltas. The availability of sediment for transport by wind is determined by its moisture content, vegetation cover, crusting and cohesion. The transport capacity of the wind is a cubic function of wind speed or surface shear stress above the transport threshold (see AEOLIAN PROCESSES). These conditions are satisfied in two main environments: (1) coastal areas with sandy beaches and onshore winds (e.g. the Atlantic coasts of north-west Europe, the Pacific north-west of North America, south-east and northeastern Australia and southern South Africa); and (2) subtropical and temperate desert areas.

Dune types

Aeolian dunes develop as a result of interactions between a granular bed (sand) and turbulent shearing flow (the atmospheric boundary layer). The resulting landforms are bedforms that are dynamically similar to those developed in subaqueous shearing flows (e.g. rivers, tidal currents). The morphology of aeolian dunes therefore reflects the characteristics of the sediment (primarily its grain size) and the wind (both the local shear stress, which determines local sand transport rates, and the long-term directional variability of the wind regime). Vegetation is a significant factor influencing the morphology of dunes in coastal dunefields as well as those in semi-arid and subhumid regions. Interactions with topographic obstacles may also result in dune formation.

Dunes occur in self-organized patterns that develop over time as the response of sand surfaces to the wind regime (especially its directional variability) and the supply of sand (Werner 1995). Development of these patterns is modulated by the effects of changes in climate and sea level on sediment supply, dune mobility and wind regime characteristics, often resulting in the formation of a series of different generations of dunes. The dune types described below represent the steady-state attractors of the aeolian transport system and can evolve from a wide range of initial conditions. The orientation of dunes with respect to the wind regime is another aspect of the self-organizing nature of the system, in which dunes are oriented to maximize the gross sand transport normal to the crest. Characteristic features of dune patterns include close correlations between the height and spacing of dunes and systematic spatial variations in dune type, orientation and sediment volume.

Despite the variety of different dune types and the multiplicity of local names that have been used to refer to them, satellite images show that dunes of essentially similar form occur in widely separated areas, and occur in five main morphologic types (Figures 50 and 51). The only dune form restricted to coastal areas is the foredune because it is an integral part of the complex of near shore processes forming the beach–dune system (Bauer and Sherman 1999). Three varieties of each dune type can occur: simple (the basic form), compound (superimposition of small dunes of the same type on larger dunes), and complex (superimposition of different dune types on the primary form, (e.g. crescentic dunes on linear dunes).

Relations between the occurrence of different dune morphological types and their wind regime environment indicate that the main control of dune type is the directional variability of the wind regime (Figure 52), which can be characterized by the ratio between the resultant (vector sum) of potential sand transport (RDP or resultant drift potential) and total potential sand transport (DP or drift potential). Sand grain size, vegetation cover, topography and sediment supply play subordinate roles in the majority of cases.

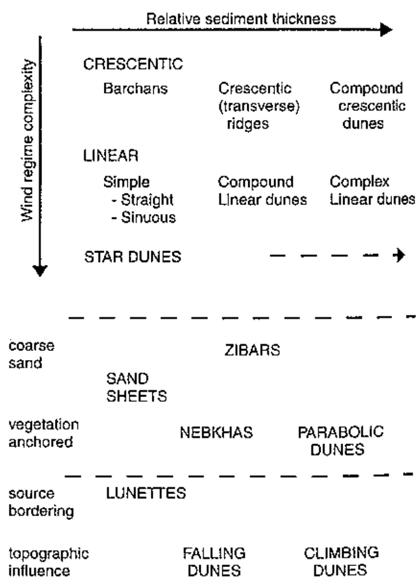

Figure 50 The main dune morphological types

The simplest dune types and patterns form in areas characterized by a narrow range of wind directions (unidirectional wind regime, RDP/DP > 0.8). In the absence of vegetation, the dominant form will be crescentic or transverse dunes with crest lines aligned approximately perpendicular to the dominant wind. Good examples are to be found in coastal areas of Namibia and the United Arab Emirates. Isolated crescentic dunes or barchans occur in areas of limited sand supply, and coalesce laterally to form crescentic or barchanoid ridges that consist of a series of connected crescents in plan view as sand availability increases. Larger forms with superimposed dunes are termed compound crescent dunes (e.g. Algodones Dunes, California; coastal areas of the Namib Sand Sea). In areas of partial vegetation cover and similar wind regimes, parabolic dunes will occur. These dunes are characterized by a U or V shape with a 'nose' of active sand and two partly vegetated arms that trail up wind. They are common in many coastal dunefields and semi-arid inland areas and often develop from localized blowouts in vegetated sand surfaces (Wolfe and David 1997). Both crescentic and parabolic dunes tend to migrate downwind, at a rate that is inversely proportional to their height.

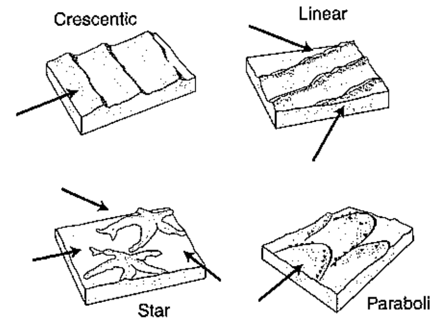

Figure 51 Schematic morphology of major dune morphological types and wind regime environments (modified from Lancaster 1995)

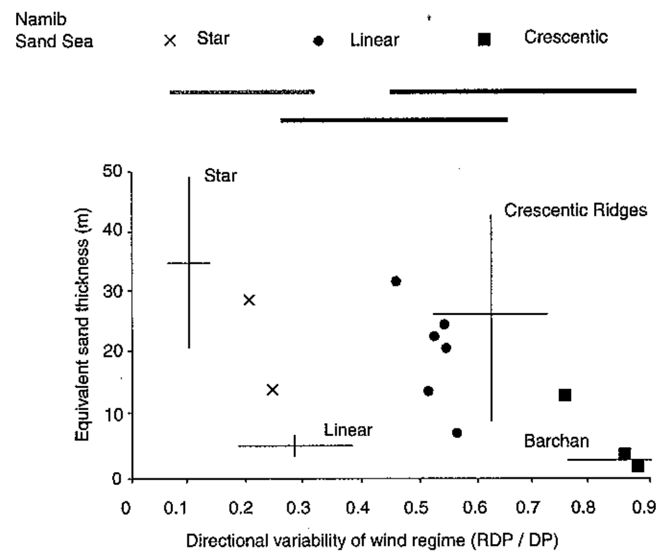

Figure 52 Relations between dune types and wind regimes. Redrawn from Wasson and Hyde (1983) with Namib Sand Sea data superimposed (symbols) and range of wind regimes for three major dune types from Fryberger (1979). Equivalent sand thickness is a measure of the sand available for dune building and represents the thickness of sand if the dunes were levelled. The directional variability of the wind regime is characterized by the ratio between resultant (RDP) and total sand drift potential (DP)

Linear dunes are characterized by their straightness, length (often more than 20 km), sinuous crestline, parallelism, and regular spacing, and high ratio of dune to interdune areas. Many linear dunes consist of a lower gently sloping plinth, often partly vegetated, and an upper crestal area where sand movement is more active. Slip faces develop on the upper part of the dune, their orientation depending on the winds of the season. The average form of the dune may be symmetrical with an approximately triangular profile, but in each season its profile tends to an asymmetric form with a concave stoss slope and a well-developed lee face. Linear dunes occur in areas of bimodal or wide unimodal wind regimes ($RDP/DP > 0.4 < 0.8$), and appear to be the most widespread dune type worldwide. Simple linear dunes occur in two forms: the straight partially vegetated dunes of the southwestern Kalahari and Simpson–Strezlecki deserts and the more sinuous 'seif'-type dunes of the Sinai and eastern Sahara. Complex linear dunes are best represented by the large (50–200-m high), widely spaced (1–2 km) linear dunes of the Namib Sand Sea.

The origins of linear dunes and their relationship to formative wind directions have been the subject of considerable controversy. A widely held view was that linear dunes form parallel to the prevailing or dominant wind direction. Their parallelism and straightness was believed to result from the existence of boundary layer roller vortices in which helicoidal flow sweeps sand from interdune areas to dunes. However, there is little empirical evidence to support such a model. Field studies of airflow and sediment transport over linear dunes (Bristow *et al.* 2000; Tsoar 1983) suggest that the fundamental mechanism for linear dune formation is the deflection of winds that approach at an oblique angle to the crest to flow parallel to the dune along its lee side and transport sand along the dune. Thus any winds from a 180° sector centred on the dune will be diverted in this manner. Linear dunes tend to extend downwind, as sinuosities in the crest migrate along its length. Evidence for lateral migration, is not conclusive.

Star dunes have a pyramidal shape, with three or four sinuous sharp-crested arms radiating from a central peak and multiple avalanche faces and are the largest dunes in many sand seas, reaching heights of more than 300 m in the eastern Namib Sand Sea and the Grand Erg Oriental in Algeria. The upper parts of many star dunes are very steep with slopes at angles of 15–30°; the lower parts consist of a broad, gently sloping (5–10°) plinth or apron. Small crescentic or reversing dunes may be superimposed on the lower flank and upper plinth areas of star dunes. Comparisons between the distribution of star dunes and wind regimes suggest that they form in multidirectional or complex wind regimes ($RDP/DP < 0.3$). A strong association between the occurrence of star dunes and topographic barriers has also been noted. Topography may modify regional wind regimes to increase their directional variability, as in the Erg Fachi Bilma or create traps for sand transport, as at Kelso Dunes and Great Sand Dunes.

The development of star dunes is strongly influenced by the high degree of form–flow interaction that occurs as a result of seasonal changes in wind direction, and the existence of a major lee-side secondary circulation. Most of the erosion and deposition involves the reworking of deposits deposited in the previous wind season. Sand, once transported to the dune, tends to stay there and add to its bulk, resulting in dunes that do not change position over time (Lancaster 1989).

Other important dune types include nebkhas or hummock dunes (common in many coastal dunefields) anchored by vegetation, lunettes (often comprised of sand-sized clay pellets) that form downwind of small playas; and a variety of topographically controlled dunes (climbing and falling dunes, echo dunes). Low relief sand surfaces such as sand sheets are common in many ergs and occupy from as little as 5 per cent of the area of the Namib Sand Sea to as much as 70 per cent of the area of Gran Desierto. Sand sheets occur where sediment availability is limited as a result of coarse sand, high water table or vegetation cover. Zibar, or low rolling dunes without slip faces composed of coarse sand, are transitional between sand sheets and crescentic dunes in some dunefields (e.g. Algodones, Skeleton Coast, Namibia).

Dune processes and dynamics

The initiation, development and equilibrium morphology of all aeolian dunes are determined by a complex series of interactions between dune morphology, airflow, vegetation cover and sediment transport rates. In turn, the developing bedforms exert a strong control on local transport rates through form–flow interactions and secondary flow circulations, leading to a dynamic equilibrium between dune morphology and local airflow. In multidirectional wind regimes, the nature of interactions between dune form and airflow change as winds vary direction seasonally, and lee-side secondary flow patterns become important in determining dune morphology and dynamics.

As dunes grow, they project into the atmospheric boundary layer so that they affect the airflow around and over them in a manner similar to isolated hills. Winds approaching the upwind toe of a dune stagnate slightly and are reduced in velocity, but likely not turbulence intensity. On the stoss, or windward slope of the dune, streamlines are compressed and winds accelerate up the slope. The degree of flow acceleration (the speed-up factor) is determined by the aspect ratio and the height of the dune. Wind speed at the crest of the dune is typically 1.1 to 2.5 times that measured immediately upwind of the dune (Figure 53a). Flow acceleration, coupled with effects of stream line curvature, on the windward slopes of dunes give rise to an exponential increase in sediment transport rates (Figure 53b) towards the dune crest (Lancaster *et al.* 1996; McKenna Neuman *et al.* 1997), resulting in erosion of the stoss slope, and a high level of erosion and deposition in crestal areas of linear and star dunes (e.g. Lancaster 1989; Livingstone 1989). Numerical models suggest that the non-linear increase in sediment transport with height on a dune limits dune size and results in an equilibrium dune configuration.

In the lee of the crest of dunes, wind velocities and transport rates decrease rapidly as a result of flow expansion between the crest and brink of the lee or avalanche face and flow separation on the avalanche face itself. There is a complex pattern of flow separation, diversion and re-attachment on the lee slopes of dunes, which is determined by the angle between the wind and the dune crest (angle of attack) and the dune aspect ratio (Walker and Nickling 2002). Secondary flows, including lee-side flow diversion, are especially important where winds approach the dune obliquely, and are an important process on linear and many star dunes.

High angles of attack on high aspect ratio (steep) dunes result in flow separation in the lee, while lower angles of attack result in flow diversion along the lee slope, whereas low aspect ratio dunes are characterized by flow expansion. Flow separation results in the development of an eddy in the lee of the dune, which may have the form of a roller vortex if flow is truly transverse, with the separation cell extending downwind for 4 to 10 times the height of crescentic dunes. When flow is oblique to the dune crest a helical vortex develops. The oblique flow is deflected along the lee slope parallel to the dune crest, with the degree of deflection being inversely proportional to the incidence angle between the crestline and the primary wind. Field studies indicate that the lee-side helical eddy affects the whole of the lee side on simple (5–10-m high) linear dunes, but extends for only 10–20 per cent of the height of large (50–150-m high) complex linear dunes and 40 m high star dunes. Changes in the local incidence angle between primary winds and a sinuous dune crest result in a spatially varying pattern of deposition and along-dune transport on the lee face. Deposition dominates where winds cross the crest line at angles approaching 90°, and erosion or along-dune transport occurs where incidence angles are <40°. Studies on flow-transverse dunes indicate that, downwind of and above the flow separation cell, there is a series of wakes that gradually expand, diffuse and mix downwind for a distance of as much as 25 to 30 times the height of the dune (Walker and Nickling 2002).

Flow separation also causes fallout of previously saltating sand grains. Field experiments show that 95 per cent of the sand transported over the crest is deposited within a metre of the crest, with the rate of deposition decreasing exponentially downwind (Nicking *et al.* 2002). High rates of deposition in the immediate lee of the crest result in oversteepening of the slope and avalanching of grains to form slip faces. All sand transported over the crest of crescentic dunes is deposited, so that they are typically 'sand trapping' bedforms. As a result, their movement can be described by: $c = Q/yh$ where c is the migration rate, Q is the bulk volumetric sand transport rate, y is the bulk density of sand and h is dune height.

Challenges and opportunities in dune studies

The past two decades have seen a dramatic change in the level of understanding of dune dynamics and morphology through intensive field studies of processes and synoptic views of sand seas provided by satellite images. As a result, the fundamentals of dune dynamics and the formative factors of major dune types are known in

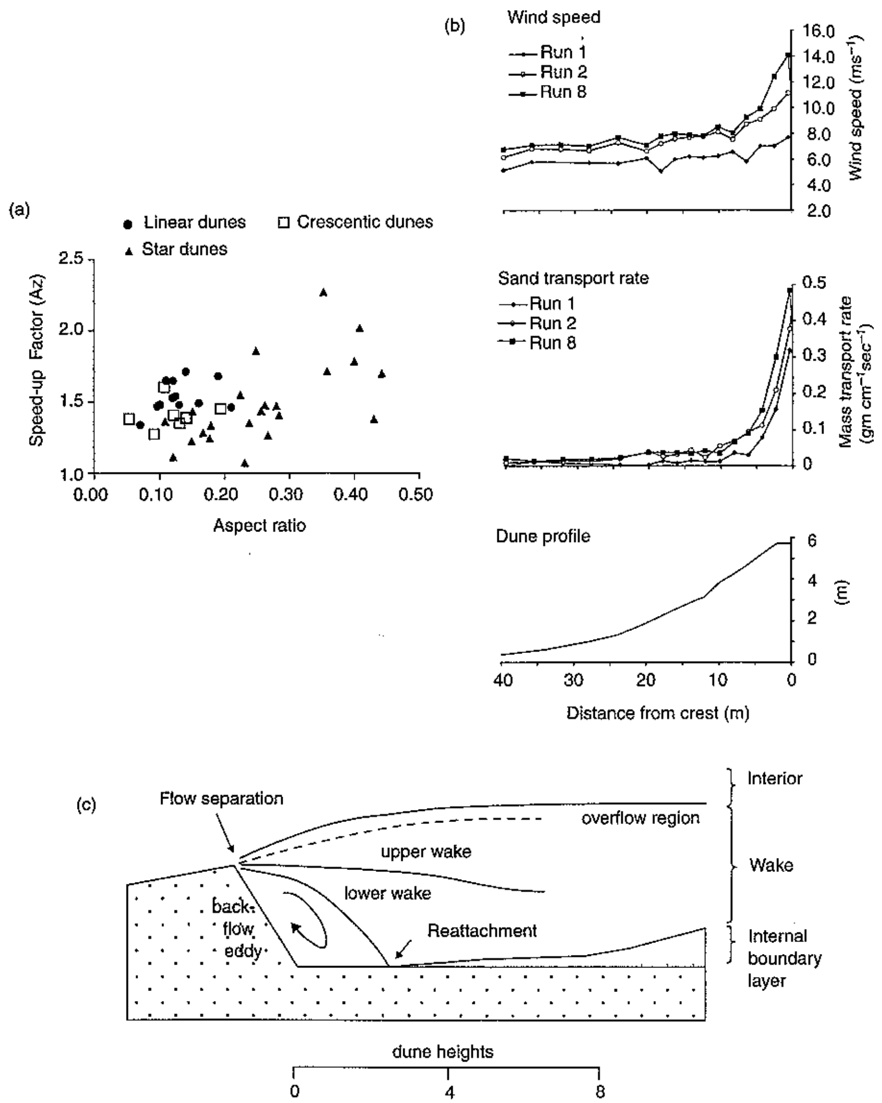

Figure 53 Elements of dune dynamics: (a): velocity speed-up (from Lancaster 1995); (b) winds and sediment transport rates on the stoss slope (from McKenna Neuman *et al.* (1997) (c) lee-side flow separation and wake mixing (from Walker and Nickling 2002)

some detail. Not well known are processes leading to dune initiation, the dynamics of lee-side processes (including avalanching), and the controls of dune size and spacing. It has also proved very difficult to extrapolate the results of short-term studies of dune processes to understanding of long-term or even annual dune dynamics. One promising approach is to develop numerical models of dune and dunefield evolution (Werner 1995). The other is to use ground-penetrating radar to image dune sedimentary structures, which provide a record of the results of dune-forming processes on a variety of timescales and allow empirical models of dune evolution to be developed. A good example of this approach is the five-stage model of linear dune development put forward by Bristow *et al.* (2000).

References

Bauer, B.O. and Sherman, D.J. (1999) Coastal dune dynamics: problems and prospects, in A.S. Goudie, I. Livingstone and S. Stokes (eds) *Aeolian Environments, Sediments and Landforms*, 71–104, Chichester, New York: Wiley.

Bristow, C.S., Bailey, S.D. and Lancaster, N. (2000) Sedimentary structure of linear sand dunes, *Nature* 406, 56–59.

Fryberger, S.G. (1979) Dune forms and wind regimes, in E.D. McKee (ed.) *A Study of Global Sand Seas*, 137–140, United States Geological Survey, Professional Paper 1,052.

Kocurek, G. and Lancaster, N. (1999) Aeolian sediment states: theory and Mojave Desert Kelso Dunefield example, *Sedimentology* 46, 505–516.

Lancaster, N. (1989) The dynamics of star dunes: an example from the Gran Desierto, Mexico, *Sedimentology* 36, 273–289.

——(1995) *Geomorphology of Desert Dunes*, London: Routledge.

Lancaster, N., Nickling, W.G., McKenna Neuman, C.K. and Wyatt, V.E. (1996) Sediment flux and airflow on the stoss slope of a barchan dune, *Geomorphology* 17, 55–62.

Livingstone, I. (1989) Monitoring surface change on a Namib linear dune, *Earth Surface Processes and Landforms* 14, 317–332.

McKenna Neuman, C., Lancaster, N. and Nickling, W.G. (1997) Relations between dune morphology, airflow, and sediment flux on reversing dunes, Silver Peak, Nevada, *Sedimentology* 44, 1,103–1,114.

Nickling, W.G., McKenna Neuman, C. and Lancaster, N. (2002) Grainfall processes in the lee of transverse dunes, Silver Peak, Nevada, *Sedimentology* 49, 191–211.

Tsoar, H. (1983) Dynamic processes acting on a longitudinal (seif) dune, *Sedimentology* 30, 567–578.

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.

Wasson, R.J. and Hyde, R. (1983) Factors determining desert dune type, *Nature* 304, 337–339.

Werner, B.T. (1995) Eolian dunes: computer simulations and attractor interpretation, *Geology* 23, 1,107–1,110.

Wolfe, S.A. and David, P.P. (1997) Parabolic dunes: examples from the Great Sand Hills, southwestern Saskatchewan, *Canadian Geographer* 41, 207–213.

Further reading

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

Nickling, W.G. and McKenna Neuman, C. (1999) Recent investigations of airflow and sediment transport over desert dunes, in A.S. Goudie, I. Livingstone and S. Stokes (eds), *Aeolian Environments, Sediments and Landforms*, Chichester: Wiley.

Pye, K. and Tsoar, H. (1990) *Aeolian Sand and Sand Dunes*, London: Unwin Hyman.

SEE ALSO: aeolian processes; barchan; draa (megadune); dune mobility

NICK LANCASTER

DUNE, COASTAL

Foredune

Foredunes are shore-parallel dune ridges formed on the top of the backshore by aeolian sand deposition within vegetation. They may range from scattered hummocks or nebkha, relatively flat terraces, to markedly convex ridges. Actively forming foredunes occupy a foremost seaward position, but not all foremost dunes are foredunes. Other dune types may occupy a foremost position on eroding coasts or coasts where foredunes are unable to form. Foredunes generally fall into two main types, incipient and established foredunes.

INCIPIENT FOREDUNE

Incipient foredunes are new, or developing foredunes forming within pioneer plant communities. They may be formed by sand deposition within discrete or relatively discrete clumps of vegetation or individual plants forming shadow dunes, hummocks or nebkha.

Incipient foredunes may also form on the backshore by relatively laterally continuous alongshore growth of pioneer plant seedlings and/or rhizomes in the wrack line or spring high tide region (Hesp 1989). Morphological development principally depends on plant density, height and cover, wind velocity and rates of sand transport (Davies 1980).

Shadow dunes, hummocks, embyro dunes and nebkha all form due to high localized drag within and behind individual plants and clumps of plants. Wind velocities experience rapid deceleration on reaching the plants, local acceleration around the plants, and flow separation behind the plants (Hesp 1999).

Relatively continuous plant canopies variously impact the wind/sand flow depending on plant density, distribution and height. High, dense canopies act to reduce flow velocities very rapidly, and sand transport (saltation and traction) is markedly reduced from the leading edge. In canopies which vary alongshore in density or distribution, foredune morphology also varies (Nickling and Davidson-Arnott 1990). Plant density is increased as wind velocities increase as the vegetation bends and streamlines to the wind.

Incipient foredunes generally display one of three morphological types: ramps, terraces and ridges. Swales (lee dune depressions) are generally created by seaward accretion of a foredune. They develop as low to limited aeolian deposition zones (Hesp 2002).

ESTABLISHED FOREDUNE

Established foredunes develop from incipient foredunes and are commonly distinguished by the growth of intermediate or 'secondary' plant species, and/or by their greater morphological complexity, height, width, age and geographical position.

Foredunes range from very low, and commonly scattered, dunes less than a metre or so in height on some barrier islands dominated by overwash and in areas of limited sediment supply to over 30 m in height in some instances. The morphological development and evolution of established foredunes depends on a number of factors including: sand supply, beach width and fetch, surfzone-beach type, the degree of vegetation cover, plant species present (a function of climate and biogeographical region), the rate of aeolian sand accretion and erosion, the frequency and magnitude of wave and wind forces, the occurrence and magnitude of storm erosion, dune scarping and overwash processes, the medium to long-term beach or barrier state (stable, accreting or eroding), and increasingly, the extent of human interference and use (Davidson-Arnott and Law 1996; Short and Hesp 1982; Hesp 1999).

The wind flow is topographically accelerated over foredunes, particularly up stoss slopes and over crests. However, the variable vegetation cover of foredunes and their topographic variability leads to local decelerations and variations in roughness length (Arens 1997), and these become more pronounced as foredune morphological complexity and vegetation cover increases.

Foredune Plain

Foredunes may gradually, or rapidly, become isolated from accretion and erosion processes by the seaward development of a new incipient foredune which itself may evolve into an established foredune. The original foredune then becomes a relict foredune as it is largely or wholly removed from a foremost beach position. Systematic beach progradation over time frames of tens to thousands of years has led to the development of wide foredune plains.

Blowout

A blowout is a saucer-, cup-, bowl or trough-shaped depression or hollow formed by wind erosion on a pre-existing sand deposit. The adjoining accumulation of sand, the depositional lobe, derived from the depression and possibly other sources, is normally considered part of the blowout (Nordstrom *et al.* 1990).

Blowout morphology may be highly variable, ranging from cigar-shaped, V-shaped, scooped hollow, and cauldron and corridor types, from pits to elongated notches, troughs or broad basins, and saucer and trough blowouts (Cooper 1967). Saucer blowouts are semicircular or saucer-shaped and often appear as shallow dishes. Deeper cup- or bowl-shaped blowouts may evolve from these. Trough blowouts are generally more elongate, with deeper deflation floors and basins, and with steeper, longer erosional lateral walls or slopes.

INITIATION

Blowouts may be initiated in a variety of ways including:

1. wave erosion of dunes followed by wind erosion of dune scarps;
2. die-back of vegetation following dune wave erosion and subsequent wind erosion;
3. wind erosion of overwash hollows and fans;
4. topographic acceleration of airflow over (or through) dunes, dune cols, scarps and cliffs;
5. where the vegetation cover is naturally low, or is weakened, reduced or dies due to a prolonged dry or arid period;
6. vegetation die-back due to soil nutrient depletion;
7. localized aridity (e.g. on dune crests) reducing plant cover;
8. the activities of animals and humans;
9. water erosion;
10. high velocity wind erosion leading to either erosion, or sand inundation and burial.

Once initiated, the subsequent morphologic development may depend on the size of the initial constriction, the height and width of the dune in which the blowout is developing, the degree and type of vegetation cover, the magnitude of regional winds, and the degree of exposure to winds from various directions (Hesp 2002; Jennings 1957).

FLOW DYNAMICS

Flow in saucer blowouts is complex with flow separation occurring around much of the erosion walls. Sand erosion and deposition are also complex as a result of varying wind speeds and directions, although, in general, deflation basins deepen in most blowouts studied. Saucer blowouts commonly grow in length upwind against the prevailing wind.

The flow up trough blowouts is commonly topographically accelerated, and displays marked single and double jets up the deflation basin, corkscrew vortices over the lateral erosional wall crests, and rapid flow deceleration, lateral expansion and flow separation over the depositional lobe. Topographic steering can be significant (Hesp and Hyde 1996).

Parabolic dune

Parabolic dunes (also termed U-dunes, upsiloidal dunes, hairpin dunes) are typically U-and V-shaped dunes characterized by short to elongate, trailing ridges which terminate downwind in U- or V-shaped depositional lobes. The depositional lobes may be simple, relatively featureless sandsheets, or textured with a variety of dune forms (e.g. transverse dunes, barchanoidal dunes, etc). Deflation basins and plains, slacks, seasonal wetlands, ponds, lagoons and gegenwalle ridges occupy the area between the trailing ridges.

INITIATION

Parabolic dunes typically evolve in a number of ways including:

1. from blowouts (Pye 1983). In many cases, the blowout depositional lobe continues to advance downwind forming trailing ridges;
2. evolution from the landward and downwind margins of transgressive sandsheets and dunefield.

Blowouts and parabolic dunes may be formed on both stable (sediment supply balanced) and accreting/prograding (positive sediment supply) coasts which experience occasional or regular high energy wind events (Hesp 2002). They are commonly formed on eroding coasts where the foredune stability is reduced by wave erosion, and subsequent wind erosion (e.g. Ruz and Allard 1994).

MORPHOLOGY

Two principle sub-types of parabolic dune are common: long-walled types and squat, elliptical types. The multiple development of these leads to there being two principle sub-types of parabolic dunefields: long-walled types and imbricate types (Trenhaile 1997).

Long-walled parabolic dunes display long trailing ridges and extensive deflation basins. They are particularly well developed on relatively flat terrain, in regions of low heath or shrubland, high sand supply and strong, more unidirectional winds. Some parabolic dunes display a squat, shorter form, often with more semicircular or elliptical deflation basins. Multiple development results in the dunes overlapping each other in an imbricate fashion. They commonly develop in wetter areas, on flat terrain where deflation depths are limited and/or wind speeds are relatively low, on steep terrain, in less unidirectional or multidirectional wind regimes and/or in dense, tall vegetation where the rate of advance is low and/or migration is impeded.

EVOLUTION

Deflation basins tend to continue to erode until a base level is reached such as the seasonally lowest water-table level, a calcrete (or other cemented/indurated) layer, an armoured surface such as a pebble, shell, pumice or artifact surface. Trailing ridges develop due to trapping of the outside, marginal lateral edge of the depositional lobe as it migrates downwind. This sediment is trapped while the inside (deflation plain) portion of the ridge is eroded. Depositional lobes are arcuate, hairpin, V-shaped, radial or parabolic-shaped depending on wind direction, lobe height and volume, vegetation cover and species type, and speed of migration.

RATES OF MIGRATION

Rates of parabolic dune advance or migration vary considerably depending on the morphology, slope and type (e.g. sandy vs rocky) of terrain the dunes are moving across, the vegetation cover and type (e.g. woodland vs grassland), wind velocities and directional variability of the wind. Dune migration rates range from 0.05 to $25\,m\,yr^{-1}$.

Transgressive dunefield and sheet

Transgressive dunefields and sheets are aeolian sand deposits formed by the downwind or alongshore movement of sand over vegetated to semi-vegetated terrain. Such sheets and dunefields may range from quite small (hundreds of metres in alongshore and landward extent) to draa or megadune size fields. They may be completely unvegetated, partially vegetated or fully vegetated (post-formation) (Nordstrom *et al.* 1990). Sheets display little or no surface dunes; dunefields are covered with a variety of superimposed dune forms. They have also been termed mobile dunes, migratory dunes, mendano and machair.

Transgressive dunefields are particularly well developed on high wind and wave energy (west and south) coasts with significant sediment supply, and in virtually all climatic regions (tropics to the arctic).

TRANSGRESSIVE DUNEFIELD TYPES

At the gross scale, transgressive dunefields may describe tabular forms (including headland bypass dunefields), buttress forms, or climbing, clifftop and falling dunefields.

INITIATION AND DEVELOPMENT

Transgressive dunefields develop for a variety of reasons. They may form, or have formed:

1 as a response to rising sea level and/or climatic change, particularly in the period, 10,000 to 7,000 years BP.
2 in regions of high alongshore and onshore sediment supply, often in high wind and wave energy environments;
3 on coasts experiencing erosion;
4 as continental shelves were exposed and/or climate changed during the Last Glacial;
5 as a response to periods of regional sea-level fall; and
6 on coasts experiencing climatic extremes such as in arid and arctic and subarctic environments, and where vegetation growth may also be limited.

TRANSGRESSIVE DUNEFIELD LANDFORMS

Active transgressive dunefields may extend (and/or migrate) directly alongshore, obliquely onshore or directly onshore. Dunefields migrating alongshore are typically characterized by transverse (and other) dunes extending inland from the backshore. Interdunes are dominated by nabkha, deflation flats, sandsheets and overwash plains and fans. Dunefields migrating obliquely or normally onshore are usually characterized by a small to extensive deflation basin (or plain) or a series of slacks on the upwind or seaward side, a small to extensive, mobile to partially vegetated sandsheet or dunefield, and a long-walled, commonly sinuous 'main' slipface or precipitation ridge on the landward side.

The surfaces of active transgressive dunefields are commonly covered with a variety of dune types (including 'desert' dune types) ranging from simple domes, transverse dunes and BARCHANS, to barchanoidal and sinuous transverse and oblique dunes, to complex aklé or network and star dune forms (e.g. Hunter *et al.* 1983).

Deflation plains and basins typically lie parallel to the shore and are eroded down to a base level such as a CALCRETE pavement, the seasonally lowest water table, older dune surfaces and PALAEOSOLs, carbonate or bedrock.

Active transgressive dunefields may display a variety of generally smaller scale dune forms and environments, including remnant knobs, hummocks, bush pockets, nabkha, shadow dunes and 'rim dune' (around the margins of washover fans) (Nordstrom *et al.* 1990).

Precipitation ridges (long-walled or main slipfaces) commonly occur along the downwind and surrounding margins of transgressive dunefields. Where the dunefields are migrating in one primary direction, they generally have one precipitation ridge. Where they are expanding/migrating landwards and alongshore, they may have two to many precipitation and trailing ridges.

References

Arens, S.M. (1997) Transport rates and volume changes in a coastal foredune on a Dutch Wadden island, *Journal of Coastal Conservation* 3, 49–56.
Cooper, W.S. (1967) *Coastal Sand Dunes of California*, Geological Society of American Memoir 101.
Davidson-Arnott, R.G.D. and Law, M.N. (1996) Measurement and prediction of long-term sediment supply to Coastal Foredunes, *Journal of Coastal Research* 12, 654–663.
Davies, J.L. (1980) *Geographical Variation in Coastal Development*, London: Longman.
Hesp, P.A. (1989) A review of biological and geomorphological processes involved in the initiation and development of incipient foredunes, in C.H. Gimmingham, W. Ritchie, B.B. Willetts and A.J. Willis (eds) *Coastal Sand Dunes*, Proceedings of the Royal Society of Edinburgh, 96B, 181–202.
——(1999) The beach backshore and beyond, in A.D. Short (ed.) *Handbook of Beach and Shoreface Morphodynamics*, 145–170, Chichester: Wiley.
——(2002) Foredunes and blowouts: initiation, geomorphology and dynamics, in P. Gares (ed.) 29th Binghamton Geomorphology Symposium: 'Coastal Geomorphology' *Geomorphology* 48 (1/3): 245–268.
Hesp, P.A. and Hyde, R. (1996) Flow dynamics and geomorphology of a trough blowout, *Sedimentology* 43, 505–525.
Hunter, R., Richmond, B.M. and Alpha, T.R (1983) Storm-controlled oblique dunes of the Oregon Coast, *Geological Society of America Bulletin* 94, 1,450–1,465.
Jennings, J.N. (1957) On the orientation of parabolic or U-dunes, *Geographical Journal* 123, 474–480.
Nickling, W.G. and Davidson-Arnott, R.G.D. (1990) Aeolian sediment transport on beaches and coastal dunes, in R.G.D. Davidson-Arnott (ed.) *Proceedings Symposium on Coastal Sand Dunes*, 1–35, Ottawa: NRC.
Nordstrom, K.F., Psuty, N.P. and Carter, R.W.G. (eds) (1990) *Coastal Dunes: Form and Process*, London: Wiley.
Pye, K. (1983) Coastal dunes, *Progress in Physical* Geography 7(4), 531–557.
Pye, K. and Tsoar, H. (1990) *Aeolian Sand and Sand Dunes*, London: Unwin Hyman.
Ruz, M.-H. and Allard, M. (1994) Coastal dune development in cold climate environments, *Physical Geography* 15(4), 372–380.
Short, A.D. and Hesp, P.A. (1982) Wave, beach and dune interactions in southeast Australia, *Marine Geology* 48, 259–284.
Trenhaille, A.S. (1997) *Coastal Dynamics and Landforms*, Oxford: Oxford University Press.

PATRICK HESP

DUNE, FLUVIAL

Dunes are the commonest sedimentary feature in sand- or silt-bedded streams. They are roughly triangular in profile with a gentle upstream slope and a steeper downstream slope (see DUNE, AEOLIAN). Dune height and wavelength are directly related to water depth. Reaching heights of up to one-third of flow depth they are commonly 0.1 m to 10 m high with a wavelength 4 to 8 times flow depth (Knighton 1998). They frequently form in streams with higher intensity flows than those with RIPPLE bedforms but, like ripples, they migrate downstream through the processes of erosion on the upstream slope followed by deposition on the downstream slope. Separation of flow from the crest of dunes generates large-scale turbulence in rivers and the downstream migration and change of dune form is an important mechanism of bedform adjustment to changing river discharge. Cross-bedding structures resulting from dune migration are often preserved in alluvial deposits and can be used to interpret palaeohydraulic processes.

Reference

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

GILES F.S. WIGGS

DUNE MOBILITY

Approximately 20 per cent of the world's drylands are covered by aeolian sands, within which desert dunes (see DUNE, AEOLIAN) occur, while coastal dunes occur in a range of climatic settings. In dryland and coastal settings dunes may be mobile, but many dunes are not mobile in the sense of the dune body migrating across the landscape. Dune mobility therefore needs to be considered in terms of dune setting, dune type, and the nature of aeolian activity upon the dune surface.

The mobility of sand dunes is in broad terms a function of the relationship between the forces of erodibility, affecting the potential of the dune surface to be eroded by the wind, and erosivity, which is the potential of the agencies effecting AEOLIAN PROCESSES to move sediment. Dune mobility may be assessed from climatological data using a dune mobility index (e.g. that of Lancaster 1988) that integrates the forces that affect erodibility, such as P/PET, and those that affect the erosivity, which relate to the wind field. Mobility can also be assessed in the field, by monitoring dune surface change and airflow (e.g. Wiggs *et al.* 1995) or dune movement in the landscape (e.g. Hastenrath 1987)

Since dunes, except those that have become lithified, are mainly comprised of unconsolidated sand-sized sediment, their potential to be moved by mobile air via processes of sediment entrainment may appear to be considerable. Reality is more complex, however, since winds have to exceed a threshold velocity for entrainment to take place, and dune bodies can, once formed, store moisture and act as a host to plants, which can markedly reduce erodibility. Dune mobility can be a discontinuous process too. Winds capable of entraining and moving sediment do not occur continuously, but vary seasonally and daily. For example, sand transport on dunes in the Namib Sand Sea generally occurs in response to fairly persistent but moderate south westerly winds during summer months, while in winter short-lived but high magnitude wind events transport sediment from an easterly direction (Livingstone 1989).

Vegetation and dune mobility

Isolated or widely spaced plants on a dune can lead to localized zones of higher wind velocities as airflow is streamlined around the obstacle (Thomas and Tsoar 1991). But in general, dunes that possess some form of surface vegetation or crusting have, all other things being equal, a lower erodibility than those that do not. Crusts and plants can play several roles in affecting the potential mobility of surface sediments (Wolfe and Nickling 1993): protection of the sediment immediately below the plant or crust, increasing surface roughness and thereby reducing wind velocity, and trapping any moving sediment grains. A partial or discontinuous vegetation cover does not totally exclude sand movement but it may anchor a dune plinth and inhibit dune migration or lateral movement. On partially vegetated dunes, different studies have identified various threshold vegetation covers, ranging from *c.*6 per cent (Marshall 1970) to 30 per cent (Ash and Wasson 1983) above which any aeolian activity ceases. However, the impact of a given cover will vary not only according to plant shape and porosity but to both ambient wind velocities and position on the dune body (Wiggs *et al.* 1995).

Dune size and mobility

All other things being equal, smaller dunes move or experience surface change more quickly than large dunes. This is because, for a given sediment transport event, the volume of sand that can be moved represents a smaller component of the total volume of sand of a large dune than of a small dune. The ability of a dune to retain its form and position as environmental conditions change has been called 'dune memory' by Warren and Kay (1987), with small dunes of low volume having little memory, and therefore adjusting relatively rapidly to wind events, while large dunes with large volumes have 'mega memories' that may record histories spanning millennia.

Mobility of different dune types

Different basic dune forms develop in different wind directional regimes (Fryberger 1979; Thomas 1997). Generally, barchans and transverse dunes form in unimodal sand transporting wind regimes, linear dunes in bimodal or wide unimodal regimes, and star dunes in multimodal regimes, where regime refers to the overall annual directional pattern of sand transporting winds.

These different regimes determine the general types of mobility or, more appropriately activity, of these dune types. Transverse dunes are mobile in the true sense of the word, since with transport for a single direction the dunes are able to migrate. Migration rates differ between and within dunefields according to the available transport energy, but given the principle of dune memory (see above), in any location larger dunes will move more slowly than small dunes, as expressed by

$$c_r = (q_c - q_i)/h\gamma_p$$

where c_r is the migration rate, q_c is the mass transport rate at the dune crest, q_i is the mass transport rate at the dune base, h is dune height and γ_p is the bulk density of the sediment. A number of studies of migration rates have been conducted in different deserts, and are summarized in Thomas (1992) with examples of rates given in Table 14.

Linear dunes are extending forms. Net sediment transport is along the dune in the resultant direction of transport generated by the combined effect of bimodal winds. This can lead to elongation of the dune at the downwind end, but also to some lateral movement if one direction has greater transport potential than the other and if the dune plinth is not anchored by vegetation. Lateral migration can be extremely slow, for example at a rate of 50–100 m over the past 10,000 years as suggested by Rubin (1990) from evidence in the Strezlecki Desert in Australia. Other studies from Namibia and the Sinai Desert suggest elongation rates may range from less than 2 m to over 14 m per annum.

Table 14 Examples of barchan and transverse dune migration rates

Location	Dune height (m)	Migration rate (m yr^{-1})
Barchan dunes, southern Peru	1	32
	7	9
Barchan dunes, Salton Sand Sea, California	3.1	27
	8.2	14
Transverse dunes, Erg Oriental, Saudi Arabia	35	0.3
	240	0.16

Source: Data from various authors

Star dunes can be regarded as sand accumulating forms that, under the interactive effect of sand transport from at least three directions, gain in volume and height over time. The individual arms of the dune may, on a seasonal basis, behave as if they are transverse or linear forms and display displacements of up to 20 m (Lancaster 1989). If any of the contributory transport directions has a net advantage over the others, some migration of the dune body may occur over time.

References

Ash, J.E. and Wasson, R.J. (1983) Vegetation and sand mobility in the Australian desert dunefield, *Zeitschrift für Geomorphologie Supplementband* 45, 7–25.

Fryberger, S. (1979) Dune form and wind regime, *US Geological Survey Professional Paper* 1,052, 137–169.

Hastenrath, S. (1987) The barchan sand dunes of southern Peru revisited, *Zeitschrift für Geomorphologie NF* 31, 167–178.

Lancaster, N. (1988) Development of linear dunes in the southwestern Kalahari, southern Africa, *Journal of Arid Environments* 14, 233–244.

——(1989) Star dunes, *Progress in Physical Geography* 13, 67–91.

Livingstone, I. (1989) Monitoring surface change on a Namib linear dune, *Earth Surface Processes and Landforms* 14, 317–332.

Marshall, J.K. (1970) Assessing the protective role of shrub-dominated rangeland vegetation against soil erosion by wind, *Proceedings XI International Grassland Congress*, 19–23.

Rubin, D.M. (1990) Lateral migration of linear dunes in the Strezlecki desert, Australia, *Earth Surface Processes and Landforms* 15, 1–14.

Thomas, D.S.G. (1992) Desert dune activity: concepts and significance, *Journal of Arid Environments* 22, 31–38.

——(1997) Sand seas and aeolian bedforms, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology*, 373–412, Chichester: Wiley.

Thomas, D.S.G. and Tsoar, H. (1991) The geomorphological role of vegetation in desert dune systems, in J.B. Thornes (ed.) *Vegetation and Erosion*, 471–489, Chichester: Wiley.

Warren, A. and Kay, S. (1987) The dynamics of dune network, in L. Frostick and I. Reid (eds) *Desert Dunes, Ancient and Modern, Geological Society of London Special publication* 35, 205–212, Oxford: Blackwell.

Wiggs, G.F.S., Thomas, D.S.G., Bullard, J.E. and Livingstone, I. (1995) Dune mobility and vegetation cover in the southwest Kalahari Desert, *Earth Surface Processes and Landforms* 20, 515–529.

Wolfe, S.A. and Nickling, W.G (1993) The protective role of sparse vegetation in wind erosion, *Progress in Physical Geography* 17, 50–68.

DAVID S.G. THOMAS

DUNE, SNOW

Aeolian bedform are common in snow, and include ripples, drifts, barchans, and the like (Cornish 1914).

In recent years the size and importance of various megadunes have become appreciated, particularly in eastern Antarctica. These are transverse features that are oriented perpendicular to the regional katabatic wind direction. Their amplitudes are small (*c.*4 m), but their wavelengths range from 2 to over 4 km, and megadune crests are nearly parallel and 10–100 km in length (Frezzotti *et al.* 2002).

References

Cornish, V. (1914) *Waves of Sand and Snow*, London: Fisher Unwin.

Frezzotti, M., Gandolfi, S. and Urbini, S. (2002) Snow megadunes in Antarctica: sedimentary structure and genesis, *Journal of Geophysical Research* 107, D18, 1-1–1-12.

A.S. GOUDIE

DURICRUST

The word was introduced by Woolnough (1927) who subsequently defined the term thus (Woolnough 1930: 124–125): 'The widespread chemically formed capping in Australia, resting on a thoroughly leached sub-stratum . . . The nature of the deposit varies from a mere infiltration of pre-existing surface rock, to a thick mass of relatively pure chemical precipitate'.

The mineral matter deposited from solution falls into three main groups: (1) aluminous and ferruginous; (2) siliceous; and (3) calcareous and magnesian. Woolnough believed that bedrock was an important influence on the distribution of these three types, which in effect are broadly equivalent to (1) laterites, bauxites, FERRICRETES (see Tardy 1997; Bardossy and Aleva 1990); (2) SILCRETES; (3) CALCRETES, dolocretes.

Because of subsequent work on the individual duricrust types, the *crete*-based terminology of which had been laid down by Lamplugh (1907), Goudie (1973: 5) proposed a modified definition which resulted from a synthesis of various definitions that had already been developed for the individual types, and stressed their essentially subaerial and near-surface origin and nature:

> A product of terrestrial processes within the zone of weathering in which either iron and aluminium sesquioxides (in the case of ferricretes and alcretes) or silica (in the case of silcrete) or calcium carbonate (in the case of calcrete) or other compounds in the case of magnesicrete and the like have dominantly accumulated in and/or replaced a pre-existing soil, rock, or weathered material, to give a substance which may ultimately develop into an indurated mass.

Sometimes duricrusts may incorporate characteristics of more than one type, as with the widespread calsilcretes of the Kalahari.

To understand the origin and development of these geomorphologically important materials some general considerations need to be borne in mind. First, there is the question of the sources of the materials which contribute to the make-up of duricrusts. The primary elements can be derived from at least four main sources: the weathering of bedrock and sediment, inputs from dust and precipitation, plant residues and the dissolved solids in ground water. Then these sources have to be translocated and concentrated either by lateral transfers, or by vertical movements, whether upwards (*per ascensum*) or downward (*per descensum*). Third, the transferred materials need to be precipitated, and here a very wide range of processes come into play. Among the most important of these are changes in chemical equilibria caused by evaporation, by temperature changes, by pressure changes in the soil, air and water systems, by the action of organisms and by miscellaneous changes caused by interactions of different solution types.

Models for the origin of duricrusts normally fall into one of two categories: those involving relative accumulation and those involving absolute accumulation. *Relative accumulations* owe their concentrations to the removal of more mobile components, while *absolute accumulations* owe their concentration to the addition of materials to a profile. However, as McFarlane (1983: 20) has pointed out, the utility of this subdivision depends on important scale considerations. At one extreme the accumulation is entirely relative since laterites would not exist at all were not Fe and Al less readily mobilized during rigorous chemical weathering. At the other extreme, in hand specimens even the residual laterites on interfluves show much addition of Fe, since samples are enriched absolutely in materials which originated above them in the formerly existing column of rock, consumed to provide the residuum.

Furthermore, laterites and silcretes differ in that, while ferricretes can result from either relative or absolute accumulation of iron, silcrete can only form by absolute accumulation. Weathering provides the silica and in some cases the material (a weathering profile, for example) in which the silica is deposited.

Many of the early models of duricrust formation involved vertical processes, and especially the role of capillary rise of solutions from ground water. Vertical process models of this *per ascensum* type were complemented by *per descensum* models, in which it was believed that material leached from the upper part of a profile would accumulate lower down. Some of the material to be leached downward might be added to the top of the profile in the form of inputs of dust, etc.

However, more recently appreciation of the importance of CATENAS and toposequences, and of lateral soil-water movements, has resulted in an increasing concern with lateral transfer models. For example, Stephens (1971) argued that the silcretes of inland Australia formed from silica that was leached during lateritization in the humid upland areas of the east and then transported by rivers to low relief areas lying to the west. Similarly the detrital model of calcrete formation (Goudie 1983: 115) involves the lateral transportation and redeposition of weathered fragments of calcrete, moving from plateaux surfaces to footslopes.

One slightly unusual explanation for duricrust formation is that proposed for the silcretes of parts of Australia, where, it has been suggested, overlying or adjacent basalt sheets have played a role. Even amongst those who have proposed this association there is little agreement as to whether the supposed basaltic effect has been hydrothermal alteration, contact metamorphism, a release of silica from weathered basalt, or a reduction in the migration of pore waters caused by the presence of a basaltic caprock. Some doubt, however, whether such a special mechanism is justified (e.g. Ollier 1991) for what is such a widespread phenomenon.

Another general feature of models of duricrust formation has been the appreciation of the importance of organic processes. For example, in the case of calcrete, laboratory simulations with micro-flora (Krumbein 1968), and studies of petrography which have revealed calcified organic filaments of soil fungi, algae, actinomycetes and root hairs of vascular land plants, have caused the role of organisms to be given the attention they deserve (see Goudie 1996, for a review). In the case of laterite, various organic agencies have also been mooted. Micro-organisms could contribute to both mobilization and precipitation of materials. The transition from geothite to haematite in laterite profiles could be the result of iron bacteria activity, and desilicifying bacteria could be used to remove combined silica (kaolin) from bauxite.

Several factors contribute to the geomorphological importance of duricrusts: the thickness of the profiles, the properties of the different components of the profiles (e.g. their occasional ability to harden on exposure) and the topographic situation in which duricrusts develop. Ferricrete profiles may be as much as 60 m thick. Calcrete profiles in parts of southern Africa, western Australia and the Texan High Plains may exceed 40 m, while in Zaire and Namibia maximum depths of silicification may also be of the order of 50 m.

The hard upper crusts of duricrust profiles form only a limited proportion of the total profile

thickness. Typical values for alcrete and ferricrete hardpans are 1–10 m, for calcrete 0.1–10 m (with around 0.3–0.5 m being the most common), while for silcrete values of between 1 and 5 m appear normal.

Beneath the hardpan layer duricrusts display a variety of material types. Ferricretes, for example, often have rather erodible pallid and mottled horizons grading down into more or less coherent bedrock, while calcretes may be underlain by friable nodule horizons, and silcretes by kaolinitic clays. Related to the important geomorphological role of the differences between the properties of hardpans, sub-hardpan zones, weathered bedrock and bedrock in relatively simple profiles, is the role of alternations of different layers in complex profiles.

Another general aspect of duricrusts, which is relevant to their geomorphological impact, is the speed at which they form, and the rapidity with which they may harden on exposure. Rapid formation helps to preserve otherwise relatively ephemeral landforms (e.g. dunes or alluvial terraces). Quick rates of formation tend to be associated with duricrusts that originate through absolute accumulation rather than relative accumulation.

In spite of examples of rapid formation it is nonetheless apparent that for some of the great thicknesses of profiles to have developed, a considerable span of time (10^5–10^7 years) is required, together with a degree of land-surface stability. The Pleistocene was too short and too variable in climate for many of the great duricrust surfaces to have formed, and it may be for these reasons that so many of the world's duricrusts are of Tertiary age, or even earlier.

It is also important to realize that the geomorphological influence of duricrusts will depend to a considerable degree on the stage of evolution which the feature has reached. This affects both the overall thickness, the nature of the constituents, and the degree of induration.

Duricrusts may play a role in relief inversion (Plate 37). In the case of laterite, laterite-covered valleys may become ridges or strings of mesâs flanking lower, younger valleys, and pediments may become mesas (McFarlane 1976). The relief of laterite surfaces may be modified by pseudokarstic processes so that the central areas of laterite-capped mesas may become gradually lowered. Thus the periphery stands relatively higher, giving a soup-plate form. Likewise, the tendency for some calcretes and dolocretes to form preferentially in valleys and depressions sometimes leads to inversion of relief in times of greater erosion, whether by water or wind. Examples of such inverted calcrete relief are provided by McLeod (1966).

Summerfield (1978) has also indicated that silcrete can cause relief inversion. In stage 1 of his model silcrete forms in areas subject to inundation and possibly reaches its thickest development in proximity to rivers. In stage 2 rejuvenation of drainage occurs leading to erosion and drainage inversion. Subsequent back-wearing (stage 3) creates silcrete-capped residuals. These may be highly resistant to further destruction by weathering since on a world basis silcretes have a mean silica content of around 96 per cent and may on occasion exceed 99 per cent. Silcrete residuals of Tertiary age are widespread in Europe and Britain, where they are known as sarsen stones.

The presence of duricrust profiles with marked differences in properties between hardpans and some of the more friable and fine-grained materials beneath, creates conditions that favour the formation of PSEUDOKARST produced by subsurface flushing, and in the case of calcrete, solutional effects. Cave formation and roof collapse produce karst-like forms in laterites. Calcretes, because of their high carbonate content and relative solubility, frequently show sinkhole development and pipe formation.

Many workers have used duricrusts as indicators of palaeoclimates, and in broad terms this may be acceptable. Calcretes, for example, are for the most part, though not exclusively, currently forming in semi-arid areas where annual rainfall is around 200–500 mm, so that their presence in various Tertiary sediments in western and central Europe may be used with a fair degree of certainty to infer formerly more arid conditions with an annual water deficit.

Plate 37 A laterite-capped plateau at Panchgani in the Deccan Plateau, India. The laterite acts as a caprock and has resulted from severe tropical weathering acting on Tertiary basalts

Much more controversy surrounds silcrete, however, as indicated by Summerfield (1983), with a range of inferred climatic conditions ranging from extreme arid to humid tropical. Summerfield maintains that silcrete may form under two distinct climatic regimes. He draws a distinction between 'the non-weathering profile' silcretes, which results from localized silica mobility and concentration in high pH environments under a predominantly arid and semi-arid climate, and 'the weathering profile' silcretes whose geochemical and petrographic characteristics are indicative of silicification under a much more humid climate in highly acidic, poorly drained weathering environments.

It is normally accepted that ferricretes and alcretes form under relatively humid conditions. Alternating wet and dry seasons were widely considered to be favourable if not essential to laterite genesis. In particular it was believed that seasonally alternating conditions were necessary for sesquioxide precipitation. However, as McFarlane (1976: 45) has pointed out, there is some evidence for its formation under permanently moist atmospheric conditions.

Duricrusts are widespread features, especially in low latitudes, though relict forms occur in more temperate ones. They have many geomorphological effects. Controversial is the question of their palaeoclimatic significance. They result from a complex interplay of different source materials, transfer processes and precipitation mechanisms in the surface and near-surface environment. In the past the roles of lateral translocations and organic processes have tended to be neglected.

References

Bardossy, G. and Aleva, G.J.J. (1990) *Lateritic bauxites*, Amsterdam: Elsevier.

Goudie, A.S. (1973) *Duricrusts in Tropical and Subtropical Landscapes*, Oxford: Clarendon Press.

——(1983) Calcrete, in A.S. Goudie and K. Pye (eds) *Chemical Sediments and Geomorphology*, 93–131, London: Academic Press.

——(1996) Organic agency in calcrete development, *Journal of Arid Environments* 32, 103–110.

Krumbein, W.E. (1968) Geomicrobiology and geochemistry of the 'Narilime-crust' (Israel), in G. Miller and G.M. Friedman (eds) *Recent Developments in Carbonate Sedimentology in Central Europe*, 138–147, Heidelberg: Springer Verlag.

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

McFarlane, M.J. (1976) *Laterite and Landscape*, London: Academic Press.

——(1983) Laterites, in A.S. Goudie and K. Pye (eds) *Chemical Sediments and Geomorphology*, 7–58, London: Academic Press.

McLeod, W.N. (1966) The geology and iron deposits of the Hammersley Range area, Western Australia, *Bulletin Geological Survey of Western Australia* 117.

Ollier, C.D. (1991) Aspects of silcrete formation in Australia, *Zeitschrift für Geomorphologie* 35, 151–163.

Stephens, C.G. (1971) Laterite and silcrete in Australia: a study of the genetic relationships of laterite and silcrete and their companion materials, and their collective significance in the weathered mantle, soils, relief and drainage of the Australian continent, *Geoderma* 5(1), 5–52.

Summerfield, M.A. (1978) The nature and origin of silcrete with particular reference to southern Africa. Unpublished D.Phil. Thesis, University of Oxford.

——(1983) Silcrete, in A.S. Goudie and K. Pye (eds) *Chemical Sediments and Geomorphology*, 59–91, London: Academic Press.

Tardy, Y. (1997) *Petrology of Laterites and Tropical Soils*, Rotterdam: Balkema.

Woolnough, W.G. (1927) The duricrust of Australia, *Journal and Proceedings of the Royal Society of New South Wales* 61, 24–53.

——(1930) Influence of climate and topography in the formation and distribution of products of weathering, *Geological Magazine* 67, 123–132.

A.S. GOUDIE

DUST STORM

A large volume of predominantly silt-sized sediment blown into the atmosphere by a strong wind. The definition most widely used for this type of WIND EROSION OF SOIL event is that devised by meteorologists: a dust-raising event that reduces horizontal visibility to 1,000 m or less.

The entrainment of dust from the ground surface is controlled by the nature of the soil or sediment itself, the nature of the wind, and the presence of any surface obstacles to wind flow. Dust storms can occur in any environment given appropriate conditions of bare, unconsolidated sediment and a strong turbulent wind, but they are most common in deserts and on their margins. Most geomorphologists define dust particles according to the silt/sand boundary (i.e. less than 62.5 μm). The particles that make up desert dust storms are dominated by SiO_2, probably reflecting the importance of quartz in source areas. Grain size, mineralogy and chemical composition can be used to distinguish soil dust from other types of particles in the atmosphere such as those derived from sea salt, volcanoes and smoke particles from fire.

Different terrain types vary greatly in their susceptibility to dust storm occurrence. Important determining factors include the ratio of clay-, silt- and sand-sized particles, the soil moisture content, the compaction of sediments, the presence of particle cements such as salts or organic breakdown products, and the presence of crusts or armoured surfaces (Middleton 1990). The most favourable dust-producing surfaces are areas of bare, loose and mobile sediments containing substantial amounts of sand and silt but little clay. Terrains that satisfy these conditions are most commonly found in geomorphologically active landscapes where tectonic movements, climatic changes and/or human disturbance are responsible for rapid exposure, incision and reworking of sediment formations containing dust. Important sources of dust storms are generally located in specific, relatively small desert environments (Coudé-Gaussen 1984) such as floodplains, alluvial fans, salt pans and devegetated fossil dunes.

The most important meteorological systems capable of generating dust storms are synoptic in scale, dominated by the passage of low pressure fronts with intense baroclinal gradients that are accompanied by very high-velocity winds. Such frontal passage is the dominant dust-generating mechanism in many of the world's dusty regions, including Australia, northern China and Mongolia, Central Asia, the Levant, the Mediterranean coast of north Africa, the Sahelian latitudes of west Africa, the High Plains of the USA and the plains of the Argentine Pampas. Surface cyclones themselves may sweep out gyres of dust when circulation around the low pressure becomes very intense. Other synoptic-scale dust-raising systems include winds generated in areas with steep pressure gradients, such as in the Thar Desert of India and Pakistan and the northwesterly Shamal wind that blows down the Arabian Gulf from Iraq and Kuwait. More localized dust storms occur when katabatic winds deflate mountain foot sediments, as on the northern slopes of Kopet Dag on the Iran-Turkmenistan border, or in California (the Santa Ana wind). The high Andean Altiplano of Chile, north-west Argentina and southern Bolivia experiences strong dust-raising from the upper westerlies and similar upper airflow deflates sediments from the arid Tibetan Plateau. The cold downburst wind of a dry thunderstorm, the classic Haboob of southern Sudan, is perhaps the most common meso-scale dust-raising system, which raises dust at the gust front some kilometres in advance of the towering convective clouds.

Dust transport and deposition

Globally, the amount of material mobilized in dust storms is thought to be around a billion tons a year and up to half of this comes from the Sahara, indicating the geomorphological importance of AEOLIAN PROCESSES in moulding parts of its landscape. The world's two most active dust storm source areas are both in the Sahara: the Bodélé Depression to the south of the Tibesti Mountains and an area covering eastern Mauritania, western Mali and southern Algeria (Goudie and Middleton 2001). Their importance relative to other major global dust sources is indicated in Table 15 which shows maximum mean values of an Aerosol Index (AI) that indicates the intensity of atmospheric dust content. The AI is derived from the satellite-borne Total Ozone Mapping Spectrometer (TOMS) that detects UV-absorbing aerosols in the atmosphere.

Sediments from these and other world dust storm areas are regularly transported over great distances. Saharan dust is transported along three main trajectories: westward over the North Atlantic to North and South America; northward across the Mediterranean to southern Europe and sometimes as far north as Scandinavia, and along easterly trajectories across the eastern Mediterranean to the Middle East. Dust storm material from other major deserts also follows common trajectories, many of which are highly seasonal (Plate 38). They include flows from north-east Asia across the Pacific Ocean and from Mesopotamia down the Arabian Gulf. Dense dust loadings following such trajectories can be discerned on imagery from remote sensing platforms and many techniques have been applied to deposited material in order to detect its source. These include dust mineralogy and elemental composition, scanning electron microscopy of individual grain features and the presence of pollen and foraminifera.

While in the troposphere, dust can have effects on climate through a range of possible influences. Dust outbreaks may affect air temperatures through the absorption and scattering of solar radiation and may cause ocean cooling. Dust-induced changes in atmospheric temperatures and changes in concentrations of potential condensation nuclei may also affect convectional activity and cloud formation, thereby altering rainfall amounts.

Dust aerosols influence the nutrient dynamics and biogeochemical cycling of both marine and terrestrial ecosystems. Much of the material transported over long distances is deposited over the oceans (Prospero 1996) where dust storm sediments provide a major nutrient input. Where deposited on land, dust may affect soil formation. Dust that has a high carbonate content may be a factor in the formation of calcretes and dust contributes to the formation of other desert surface coverings such as desert varnish and case hardening of rocks. Salts carried in wind-blown dusts can act as weathering agents and increase the salinity of soils and water bodies.

LOESS is by definition a wind-deposited dust with a median grain size range of 20–30 μm (Tsoar and Pye 1987) and has been estimated to cover up to 10 per cent of the world's land area. Interestingly, however, the occurrence of loess in Africa is very limited, a fact that is surprising given the Sahara's prominence as the world's largest area of contemporary dust storm activity and evidence that suggests it produced more dust during cold phases of the Pleistocene (see below).

Table 15 Maximum mean AI values for major global dust sources determined from TOMS

Location	
Bodélé Depression of central Sahara	>30
West Sahara in Mali and Mauritania	>24
Arabia (southern Oman/Saudi border)	>21
Eastern Sahara (Libya)	>15
South-west Asia (Makran coast)	>12
Taklamakan/Tarim basin	>11
Etosha Pan (Namibia)	>11
Lake Eyre basin	>11
Mkgadikgadi basin (Botswana)	>8
Salar de Uyuni (Bolivia)	>7
Great Basin of the USA	>5

Source: Goudie and Middleton (2001)

Plate 38 A dust raising event at Disi, south-east Jordan, with dust being raised from a dry playa surface

Changing frequencies of dust storms

There is considerable evidence that dust storm frequencies can change substantially in response to climatic changes both in the long term (e.g. during the Last Glacial Maximum) and in the short term (e.g. in response to the North Atlantic Oscillation and to drought phases). Analysis of dust in cores taken from deep-sea sediments, ice caps and loess deposits has enabled the reconstruction of long-term changes in dust storm activity. Dust in North Atlantic sediments has been dated back to the early Cretaceous, although aeolian activity in the Sahara appears to have become more active in the late Tertiary and high dust loadings were a particular feature of the Pleistocene in many parts of the world.

Intensification of dust storm activity during glacial periods, such as the Last Glacial Maximum, was probably due in part to lower precipitation, although changes in wind regimes may also have contributed. It has also been suggested that increased atmospheric dust during the Last Glacial Maximum was not only a response to climate change but also a contributory factor to the change, and regional dust loadings are being built into models of climate (Mahowald *et al.* 1999).

Drought is commonly associated with an increase in dust-raising activity as vegetation cover dies off and soils dry out (Brooks and Legrand 2000). In the more recent past, the effects of drought on dust storm activity have sometimes been exacerbated by human influences on the wind erosion system. Human activity has been shown to affect dust storm activity by destabilizing soil surfaces and altering vegetation cover. The most common human impact in this respect is agriculture. Type examples of large areas in semi-arid climates converted from grasslands to cultivation subsequently becoming enhanced dust-producing regions include the Great Plains of the USA in the 1930s, the so-called Dust Bowl, and the Virgin Lands scheme of the former Soviet Union in the 1950s. Other human activities that affect changes in dust storm frequencies through the breakup of wind-resistant surfaces and/or the removal of protective vegetation cover from soils include drainage, construction, vehicle use and military movements. The relationship between dust-raising, environmental change and human impacts has meant that changes in dust storm frequency have been studied as an indicator of DESERTIFICATION.

Dust storm hazards

Airborne dust presents a variety of problems to inhabitants of desert areas. In areas where dust is raised from agricultural fields it represents a serious form of wind erosion while blowing dust and sand can cause considerable damage to crops and natural vegetation by abrasion, which is particularly critical for young shoots when fields are poorly protected by vegetation cover. The reduction in visibility caused by dust storms is a serious hazard to aviation and road transport.

Dust storms are a form of atmospheric pollution and may transmit diseases that affect plants, animals and humans. Fungus carried in Saharan dust has been implicated in disease outbreaks in coral reefs throughout the Caribbean (Smith *et al.* 1996). Micro-organisms blown in dust may settle on the skin, be swallowed or inhaled into respiratory passages. In Arizona, Valley Fever is caused by *Coccidioides immitis*, a common airborne fungus blown by dust storms. Inhalation of fine particles can also aggravate diseases such as asthma, bronchitis and emphysema. These risks to human and ecosystem health have been noted both in dust source areas and in areas of deposition after long-range transport (Griffin *et al.* 2001).

Applied geomorphologists have played a useful role in identifying dust sources and methods for preventing dust entrainment in arid zones. Jones *et al.* (1986) have suggested a general procedure for the assessment of dust hazards in urban areas after their work in the Middle East.

References

Brooks, N. and Legrand, M. (2000) Dust variability over northern Africa and rainfall in the Sahel, in S.J. McLaren and D. Kniverton (eds) *Linking Land*

Surface Change to Climate Change, 1–25, Dordrecht: Kluwer.
Coudé-Gaussen, G. (1984) Le cycle des poussières éoliennes désertiques actuelles et la sédimentation des loess peridésertiques quaternaires, *Bulletin Centre Recherche et Exploration-Production Elf-Aquitaine* 8, 167–182.
Goudie, A.S. and Middleton, N.J. (2001) Saharan dust storms: nature and consequences, *Earth-Science Reviews* 56, 179–204.
Griffin, D.W., Garrison, V.H., Herman, J.R. and Shinn, E.A. (2001) African desert dust in the Caribbean atmosphere: microbiology and public health, *Aerobiologia* 17, 203–213.
Jones, D.K.C., Cooke, R.U. and Warren, A. (1986) Geomorphological investigation, for engineering purposes, of blowing sand and dust hazard, *Quarterly Journal of Engineering Geology* 19, 251–270.
Mahowald, N., Kohfield, K., Hansson, M., Balkanski, Y., Harrison, S.P., Prentice, I.C., Schulz, M. and Rodhe, H. (1999) Dust sources and deposition during the last glacial maximum and current climate: a comparison of model results with paleodata from ice cores and marine sediments, *Journal of Geophysical Research* 104(D13), 15,895–15,916.
Middleton, N.J. (1990) Wind erosion and dust storm prevention, in A.S. Goudie (ed.) *Desert Reclamation*, 87–108, Chichester: Wiley.
Prospero, J.M. (1996) The atmospheric transport of particles to the ocean, in V. Ittekkot, P. Schafer, S. Honjo and P.J. Depetris (eds) *Particle Flux in the Ocean*, 19–52, Chichester: Wiley.
Smith, G.T., Ives, L.D., Nagelkerken, I.A. and Ritchie, K.B. (1996) Caribbean sea fan mortalities, *Nature* 383, 487.
Tsoar, H. and Pye, K. (1987) Dust transport and the question of desert loess formation, *Sedimentology* 34, 139–153.

Further reading

Morales, C. (ed.) (1979) *Saharan Dust*, Chichester: Wiley.
Péwé, T.L. (ed.) (1981) *Desert Dust*, Geological Society of America Special Paper 186.
Pye, K. (1987) *Aeolian Dust and Dust Deposits*, London: Academic Press.

NICHOLAS MIDDLETON

DYE TRACING

A large variety of dyes and related compounds have been used in water tracing and these are generally classified into non-fluorescent and fluorescent dyes, although strictly some of the substances commonly included in the latter category are not dyes but dye intermediaries or optical brightening agents (OBAs). The earliest scientific water-tracing experiments used simple colouring agents which were detected visually. Later it was discovered that some dyes, such as Rhodamine, may be detected on treated cotton hanks. These visual dyes have no advantages and several disadvantages. Visual detection requires high concentrations and observers and of the two dyes which can be absorbed onto cotton hanks, Malachite Green and Rhodamine B, the latter has been shown to be toxic. Hence, they cannot be recommended and are no longer used.

Fluorescent dyes differ from simple colouring agents in that when irradiated at a particular wavelength they emit light at a different, longer, wavelength. Hence, they can be detected in water samples in concentrations invisible to the naked eye; theoretically down to $ng\,l^{-1}$. They have become the most widely used water-tracing substances in limestone areas and have also been used successfully to trace water flow through other fractured rocks, through soils and through peat pipes. There are many fluorescent dyes and they are generally divided into three groups on the basis of their fluorescence spectra: Blue (e.g. Amino G Acid and Optical Brighteners such as Leucophor BS and Tinopal CBS-X and ABP); Green (e.g. Fluorescein (and its disodium salt Uranine), Pyranine and Lissamine) and Orange (e.g. Rhodamine WT, Sulpho-Rhodamine B and Eosine).

Field determination of the green and orange dyes is possible using a portable fluorimeter and in the laboratory a modern scanning spectrofluorimeter can distinguish between different dyes allowing multiple tracing experiments to be undertaken. A further advantage is that several of the green and orange dyes are absorbed by charcoal grains and may be released in the laboratory by an alkaline-alcohol elutant. This is particularly useful if it is necessary to monitor several sites as charcoal bags (variously known as fluocapteurs or receptors) can be deployed at each site and left for up to a week to scavenge dye. Unfortunately the charcoal also scavenges other organic substances which can make interpretation difficult. Treated cotton detectors can be used as fluocapteurs for optical brightening agents.

There are also disadvantages associated with each individual dye. Blue dyes, and especially OBAs suffer from high and very variable background and break down in sunlight; the green dyes fluoresce in the same area as certain organisms and organic substances; certain reds are potentially carcinogenic; and green, and to a lesser extent red, dyes may be lost on sediment.

Further reading

http://www.dyetracing.com. This website provides useful information on dye tracing.
Ward, R.S., Williams, A.T., Barker, J.A., Brewerton, L.J. and Gale, I.N. (1998) Groundwater tracer tests: a review and guidelines for their use in British aquifers, British Geological Survey Report WD/98/19.

JOHN GUNN

DYKE (DIKE) SWARM

A dyke (dike) is a tabular igneous body that was sub-vertical at the time of emplacement.[1] A dyke swarm is a set of coeval dykes which typically display a linear, radiating or arcuate geometry. Dyke swarm compositions range from ultramafic to felsic. The largest swarms are of basaltic composition (with diabasic texture) and are most prominent in basement terranes. Individual diabase dykes can range in width from centimetres to hundreds of metres and in length from metres to 2,000 km or more.

Radiating swarms with radii of about 100 km are associated with individual volcanic centres. Radiating swarms with radii >300 km (referred to as giant radiating dyke swarms) form the feeder systems for large igneous provinces, and are thought to focus above the centres of mantle plumes (Ernst and Buchan 2001: chapters 12 and 19). Magma can be transported both vertically and laterally in dykes. In particular, giant radiating swarms can transport magma laterally more than 2,000 km from the plume centre and can feed sills and volcanic rocks at any distance along their extent. A classic example of a giant radiating swarm is the 1267 Ma Mackenzie swarm of the northern Canadian Shield which fans over an arc of 90° and extends 2,300 km from the focal point (Fahrig 1987). There are also numerous giant radiating swarms on Venus and Mars (Grosfils and Head 1994; Ernst *et al.* 2001).

Many giant linear swarms on Earth may be fragments of giant radiating swarms, which have been dismembered during episodes of continental breakup. However, linear swarms can also be associated with spreading ridges, ophiolite complexes and rift zones. Arcuate portions of otherwise linear or radiating swarms may reflect primary geometry (i.e. changes in the regional stress field) or later deformation. In addition, some arcuate swams occur as a set of ring dykes generated above an intrusion.

In addition to their magmatic significance, dykes can also act as a barrier to groundwater flow, and may localize hydrothermal fluids along their margins. They often weather positively as linear ridges or negatively as troughs due to differential erosion.

Note

1 This differs from the traditional definition of dyke which would allow an originally horizontal sheet to be termed a dyke if it is discordant to the bedding or foliation of its host rocks. Because vertical and horizontal sheets imply fundamentally different stress conditions, originally sub-vertical sheets should be termed dykes and originally sub-horizontal sheets should be referred to as sills, regardless of the degree of discordance. Tabular bodies for which the original orientation cannot be determined or where the dip is intermediate would be termed 'sheets'.

References

Ernst, R.E. and Buchan, K.L. (eds) (2001) *Mantle Plumes: Their Identification Through Time*, Geological Society of America Special Paper 352.
Ernst, R.E., Grosfils, E.B. and Mège, D. (2001) Giant dike swarms: Earth, Venus and Mars, *Annual Review of Earth and Planetary Sciences* 29, 489–534.
Fahrig, W.F. (1987) The tectonic settings of continental mafic dyke swarms: failed arm and early passive margin, in H.C. Halls and W.F. Fahrig (eds) *Mafic Dyke Swarms*, Geological Association of Canada Special Paper 34, 331–348.
Grosfils, E.B. and Head, J.W. (1994) The global distribution of giant radiating dike swarms on Venus: implications for the global stress state, *Geophysical Research Letters* 21, 701–704.

RICHARD E. ERNST AND KENNETH L. BUCHAN

DYNAMIC EQUILIBRIUM

This is a concept which describes a situation of relatively restricted fluctuations about a mean value, together with non-stationarity of that mean. It has been propounded, especially by the American geomorphologist S.A. Schumm (1977, 1991) as a device which permits the short-term variability of 'GRADED TIME' to be viewed alongside the longer trajectory of 'CYCLIC TIME' (most especially for changes in the relief of valleys). In this sense, Schumm, in particular, has sought to bring together the apparently conflicting timescales and explanatory modes which late twentieth-century geomorphologists have often considered to be favoured by G.K. Gilbert and W.M. Davis respectively. The need to reconcile such apparent conflict may be linked to the extremely influential paper by R.J. Chorley (1960) that advocated the shift from an historical emphasis on the long-term and largely predictable progressive change in landforms represented by the Davisian cycle, towards a focus upon the dynamic and process-oriented approach associated with Gilbert and considered to be strongly encouraged by the adoption of systems thinking (see SYSTEMS IN GEOMORPHOLOGY).

At the root of the use of the concept of dynamic equilibrium in late twentieth-century geomorphology is, undoubtedly, Gilbert's 1877 development of the concept of grade (see GRADE, CONCEPT OF), which was then – interestingly – taken up by Davis who incorporated it within his cyclic models (see Chorley *et al.* 1964; Chorley *et al.* 1973). Grade itself has proved one of the thorniest and most elusive geomorphological concepts (cf. Kesseli 1941) and one which, by the year 2000, had more or less vanished from the literature.

Although ideas of equilibria at a variety of timescales and with a variety of theoretical underpinnings characterized much of Anglo-American geomorphology in the latter half of the twentieth century, it became increasingly clear that the concepts work best in very closely defined circumstances, where the physics or chemistry of the situation is relatively unobscured by historically contingent variability which is poorly susceptible to mathematical treatment (cf. Selby's 'strength EQUILIBRIUM SLOPES'). The whole notion of geomorphological equilibria was extensively reviewed by Thorn and Welford (1994; and discussion) and it must remain a matter of opinion whether one accepts their conclusion that the concepts are of central, significant and continuing explanatory value to geomorphology as a whole.

Certainly Schumm's belief in the utility of the notion of dynamic equilibrium has been persistent and pervasive. However, in the light of his equal emphasis on the role of geomorphic thresholds (see THRESHOLD, GEOMORPHIC), it might be argued that a better concept to link the emphases on process and on evolution is his 'Model 2' (Schumm 1977: 12) in which the dynamic equilibrium of cyclic time is replaced, in the same time frame, by dynamic *metastable* equilibrium. Whereas Schumm's dynamic equilibrium (*sensu stricto*) couples a general reduction in relief (that is, a non-stationary mean elevation) with oscillating episodes of cut-and-fill; his dynamic metastable equilibrium assumes long periods of effective stationarity of mean elevation, with variable erosion and aggradation interrupted by abrupt, episodic erosion. This condition would certainly better describe the kind of situation observed in Piceance Creek, Colorado (Schumm 1977: 78–81) where gullying was shown to be episodic. It would also accord more neatly with developing ideas of complex, non-linear models.

It remains the case, however, that both forms of dynamic equilibrium are difficult to identify unambiguously and, further, that neither may be said to add true clarity to our understanding of geomorphic process and form. One central problem is the uncertainty of the temporal and spatial scales at which the 'graded' gives way to the 'cyclic'. Such basic questions of the identification of the crucial timescales and spatial scales of landscape development and the inherent problems of piecing together events on different scales (cf. Schumm and Lichty 1965) remain central (see Church 1996). But whether the concepts of dynamic or dynamic metastable equilibrium are what is needed as the framework to reconcile the process study with the evolutionary one, is far from certain.

References

Chorley, R.J. (1960) Geomorphology and General Systems Theory, *US Geological Survey Professional Paper* 500-B.
Chorley, R.J., Beckinsale, R.P. and Dunn, A.J. (1973) *The History of the Study of Landforms*, vol. 2, London: Methuen.
Chorley, R.J., Dunn, A.J. and Beckinsale, R.P. (1964) *The History of the Study of Landforms*, vol. 1, London: Methuen.
Church, M.A. (1996) Space, time and the mountain – how do we order what we see? in B.L Rhoads and C.E. Thorn (eds) *The Scientific Nature of Geomorphology*, 147–170, Chichester: Wiley.
Kesseli, J.E. (1941) The concept of the graded river, *Journal of Geology* 49, 561–588.
Schumm, S.A. (1977) *The Fluvial System*, Chichester: Wiley.
——(1991) *To Interpret the Earth: Ten Ways to be Wrong*, Cambridge: Cambridge University Press.
Schumm, S.A. and Lichty, R.W. (1965) Time, space and causality in geomorphology, *American Journal of Science* 263, 110–119.
Thorn, C.E. and Welford, M.R. (1994) The equilibrium concept in geomorphology, *Annals of the Association of American Geographers* 84, 666–696 and discussion, 697–709.

Further reading

Stoddart, D.R. (ed.) (1997) *Process and Form in Geomorphology*, London: Routledge.
Summerfield, M.A. (1991) *Global Geomorphology*, Harlow: Longman.

SEE ALSO: complex response; complexity in geomorphology; cycle of erosion; denudation chronology; equilibrium shoreline; geomorphic evolution; punctuated aggradation

BARBARA A. KENNEDY

DYNAMIC GEOMORPHOLOGY

Dynamic geomorphology is defined as an emphasis in geomorphology which treats geomorphic processes as 'gravitational or molecular shear stresses, acting on elastic, plastic or fluid earth materials to produce the characteristic varieties of strain or failure which we recognize as the processes of weathering, erosion, transportation and deposition' (Strahler 1952). This emphasis was first thoroughly exemplified in the work of G.K. Gilbert (1877, 1909, 1914, 1917) and emulated by Brigadier Bagnold (1941) but was largely overlooked by geomorphologists until Horton (1945), Strahler (1952) and Tricart's initiation of the *Revue de Géomorphologie Dynamique* (1950). It is no exaggeration to say that this emphasis is today dominant in Anglo-American and Japanese geomorphology and is often equated with process geomorphology.

The work of G.K. Gilbert is the first seminal antecedent to the study of geomorphic process or dynamic geomorphology. His report on the geology of the Henry Mountains (1877) described the physical erosional processes and derived a system of laws governing progress from initial to adjusted forms; his discussion of the convexity of hill-tops introduced the role of soil creep as a dynamic process; his discussion of transportation of debris by running water was based on flume experimental data; and his paper on hydraulic mining debris in the Sierra Nevada was path-breaking in its recognition of the effects of the passage of a slug of sediment passing through the Sacramento River system over a period of sixty years, from the time of the commencement of gold mining to the time of his publication in 1917. Gilbert, apparently single-handedly, established the paradigm of dynamic or process geomorphology.

Brigadier Bagnold published his monumental *Physics of Blown Sand and Desert Dunes* in 1941. This book remains a sourcebook for students of AEOLIAN PROCESSES. But Bagnold also contributed to the fundamental understanding of beach formation processes and fluvial processes, much of this understanding deriving from his home made 2-m flume which he maintained in his home grounds. The third antecedent of contemporary dynamic geomorphology was the Uppsala school of physical geography in Sweden, where Hjulstrom (1935), Sundborg (1956) and Rapp (1960) transformed our understanding of fluvial process and drainage basin geomorphology. A fourth antecedent is surely the French collaboration of Tricart and Cailleux and the significant influence of Tricart in maintaining the momentum of a journal dedicated to dynamic geomorphology. Tricart's department of applied geomorphology at the University of Strasbourg was unique in continental Europe in its pioneering of this emphasis.

The essential vision of what constituted dynamic geomorphology was developed by Strahler (1952). Shear stresses affecting Earth materials were divided into two major categories: gravitational and molecular. Gravitational stresses activate all downslope movements of matter, hence include all mass movements, all fluvial and glacial processes. Indirect gravitational stresses activate tide- and wave-induced currents and winds. Phenomena of gravitational shear stresses were subdivided according to behaviour of rock, soil, ice, water and air as elastic or plastic solids and viscous fluids.

Molecular stresses are those induced by temperature changes, crystallization and melting, absorption and desiccation, or osmosis. These stresses act in random or unrelated directions with respect to gravity. Surficial creep results from a combination of gravitational and molecular stresses on a slope. Chemical processes of solution and acid reaction were considered separately. Strahler went on to say that a fully dynamic approach requires analysis of geomorphic processes in terms of open systems which tend to achieve steady states of operation and are self-regulatory to a large degree. Finally, he specified that formulation of mathematical models, both by rational deduction and empirical analysis of observational data, to relate energy, mass and time was the ultimate goal of the dynamic approach.

Strahler's motivation was to counteract the heavy emphasis on descriptive, deductive studies of landform development and regional geomorphologies that had come to dominate the subject in the early part of the twentieth century. This dynamic emphasis in geomorphology has also been characterized as functional geomorphology (Chorley 1978) and is contrasted with historical geomorphology.

Strahler (1992: 72–73) describes his encounter with open systems theory (Von Bertalanffy 1950) in the following way: 'It was as if a closed door had opened before me, revealing an entirely new and powerful epistemology of science – a paradigm capable of unifying all dynamic processes and forms that can be observed in the universe.' Strahler (1980) describes the five levels of systems organization which compose his mature reflections on dynamic geomorphology. Level 1, which corresponds closely with his 1952 discussion, concerns the collection of data which are considered potentially useful in understanding the geomorphic system. The data must be quantitative and must be in the fundamental dimensions of mass, length, time, temperature and their products. The system variables are grouped into (a) dynamic, (b) mass-flow, (c) geometry and (d) material property variables. The dynamic variables relate to energy, force and stress; the mass-flow variables express rates of flow of matter; the geometry variables describe size and form, and material property variables include environmental constants and regulators. The second level of analysis relates to morphological elements; the third level examines flow systems of interconnected pathways of transport of energy and matter; the fourth level describes process-form systems, characterized by self-regulation through physical feedback loops; and the fifth level, systems regulated by cybernetic feedback, links natural systems with those regulated and/or disturbed by human intervention. The agenda described is reminiscent of the Chorley and Kennedy (1971) agenda and underlines the close relation between dynamic geomorphology and the general systems framework.

Following Strahler's most important impetus to the development of dynamic geomorphology, the contributions of Schumm, Leopold, Wolman, Gregory and Walling can be seen as setting the seal on the paradigm, especially in the context of fluvial and watershed geomorphology. John Miller, the third of the triumvirate of Leopold and Miller, would surely have had a major influence, perhaps even the greatest influence, had he not died tragically at the age of 39. The reason for such a bold suggestion is that John Miller alone among this group of leaders was a geochemist as well as a geomorphologist and he was attempting to link weathering processes as well as mechanical erosional processes to drainage basin evolution. The themes of dynamic equilibrium, magnitude and frequency (see MAGNITUDE–FREQUENCY CONCEPT) of operation of geomorphic processes, and a strong bias towards fluvial process were to become the hallmark of dynamic geomorphology in the Anglo-American literature. The paradigm became hugely popular partly because of its quantitative rigour, partly because it seemed to provide specific answers in a field where thoughtful arm-waving had become a tradition and partly because it recognized the value of the combination of theory, experiment and practice.

Dynamic geomorphology has now become equated with 'process geomorphology'. A current textbook on process geomorphology notes that valid interpretations of geomorphic history must be based on a thorough understanding of the processes involved in landform development. Geomorphologists therefore must be cognizant of process mechanics prior to analysing how landform history manifests past climatic or tectonic phenomena (Ritter *et al.* 1995). Five basic principles of process geomorphology according to Ritter *et al.* are: (1) a delicate balance or equilibrium exists between landforms and processes. The character of this balance is revealed by considering both landforms and processes as systems or parts of systems; (2) the perceived balance between process and form is created by the interaction of energy, force and resistance; (3) changes in driving force and/or resistance may stress the system beyond the defined limits of stability. When these limits of equilibrium or thresholds are exceeded, the system is temporarily in disequilibrium and a major response may occur. The system will develop a different equilibrium condition adjusted to the new force or resistance controls, but it may establish the new balance in a complex manner; (4) various processes are linked in such a way that the effect of one process may initiate the action of another; and (5) geomorphic analyses can be made over a variety of time intervals. In process studies the time framework utilized has a direct bearing on what conclusions can be made regarding the relation between process and form. Therefore the time framework should be determined by what type of geomorphic analysis is desired.

These principles are clearly related to the earlier dynamic geomorphology of Gilbert, Bagnold, Hjulstrom, Sundborg, Rapp, Cailleux, Tricart and Strahler through the ideas of 'a balanced condition', the centrality of energy, force and resistance, the language of systems theory, complex response and the importance of timescale of study. One implication of the reductionist functionalism of dynamic geomorphology was the emergence of semi-independent geomorphic process schools, such that coastal, slope, glacial, periglacial, karst, aeolian and fluvial geomorphology became more formally differentiated. The demands of learning the mechanics and dynamics of process led to an isolation of the new dynamic geomorphology from those who were more interested in the evolution of landscapes over geological time. The central conundrum was articulated by Church (1980) when he commented that contemporary records of geomorphological processes were not likely to represent long-term behaviour sufficiently well to provide any firm basis for understanding landscape evolution.

Schumm and Lichty (1965) made a significant contribution to the linking of short-term and long-term studies by explicitly recognizing the different status of process variables over cyclic, graded and steady-state timescales. Steady-state timescales were appropriate for process studies; cyclic timescales would be appropriate to geological evolutionary studies and the graded timescale would be appropriate to, perhaps, the Holocene timescale. They suggested a reconciliation between timeless and time-bound aspects of geomorphology by noting that the distinction between cause and effect among geomorphic variables varies with the size of the landform/landscape under consideration as well as with time. Theirs is an interesting and valuable insight, but the problem would seem to arise from an insistence on the idea of balance or equilibrium at the core of dynamic geomorphology. In spite of the power of the dynamic geomorphology paradigm, there remains a tension between the way in which time and space scales of variability are treated and the central assumption of balance and equilibrium. The question at issue is whether the landscape is fundamentally in equilibrium or whether it is fundamentally in a transient state between equilibria which are rarely if ever achieved.

Dynamic geomorphology has revived geomorphology from its pre-Second World War slumber, has connected geomorphology with the other natural sciences of physics, chemistry and biology, and has opened up opportunities for professional accreditation alongside engineers. At the same time, it can be suggested that process geomorphology has, at least for a few decades, lost touch with both traditional geology and geography. With geology in that the diastrophic framework supplied by global plate tectonics has been difficult to marry with local-scale process studies; with geography in that the challenges of global environmental change and the role of human society have been equally difficult to marry with site-scale process studies.

References

Bagnold, R.A. (1941) *The Physics of Blown Sand and Desert Dunes*, London: Methuen.

Chorley, R.J. (1978) Bases for theory in geomorphology, in C. Embleton, D. Brunsden and D.K.C. Jones (eds) *Geomorphology: Present Problems and Future Prospects*, Oxford: Oxford University Press.

Chorley, R.J. and Kennedy, B.A. (1971) *Physical Geography: A Systems Approach*, London: Prentice Hall.

Church, M. (1980) Records of recent geomorphological events, in R.A. Cullingford, D.A. Davidson and J. Lewin (eds) *Time Scales in Geomorphology*, 13–30, New York: Wiley.

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

——(1909) The convexity of hill tops, *Journal of Geology* 17, 344–350.

——(1914) The transportation of debris by running water, *US Geological Survey Professional Paper* 86.

——(1917) Hydraulic mining debris in the Sierra Nevada, *US Geological Survey Professional Paper* 105.

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.

Ritter, D.F., Miller, J.R., Grizel, Y. and Wells, S.G. (1995) Reconciling the roles of tectonics and climate in Quaternary alluvial fan evolution, *Geology* 23, 245–248.

Schumm, S.A. and Lichty, R.W. (1965) Time, space and causality in geomorphology, *American Journal of Science* 263, 110–119.