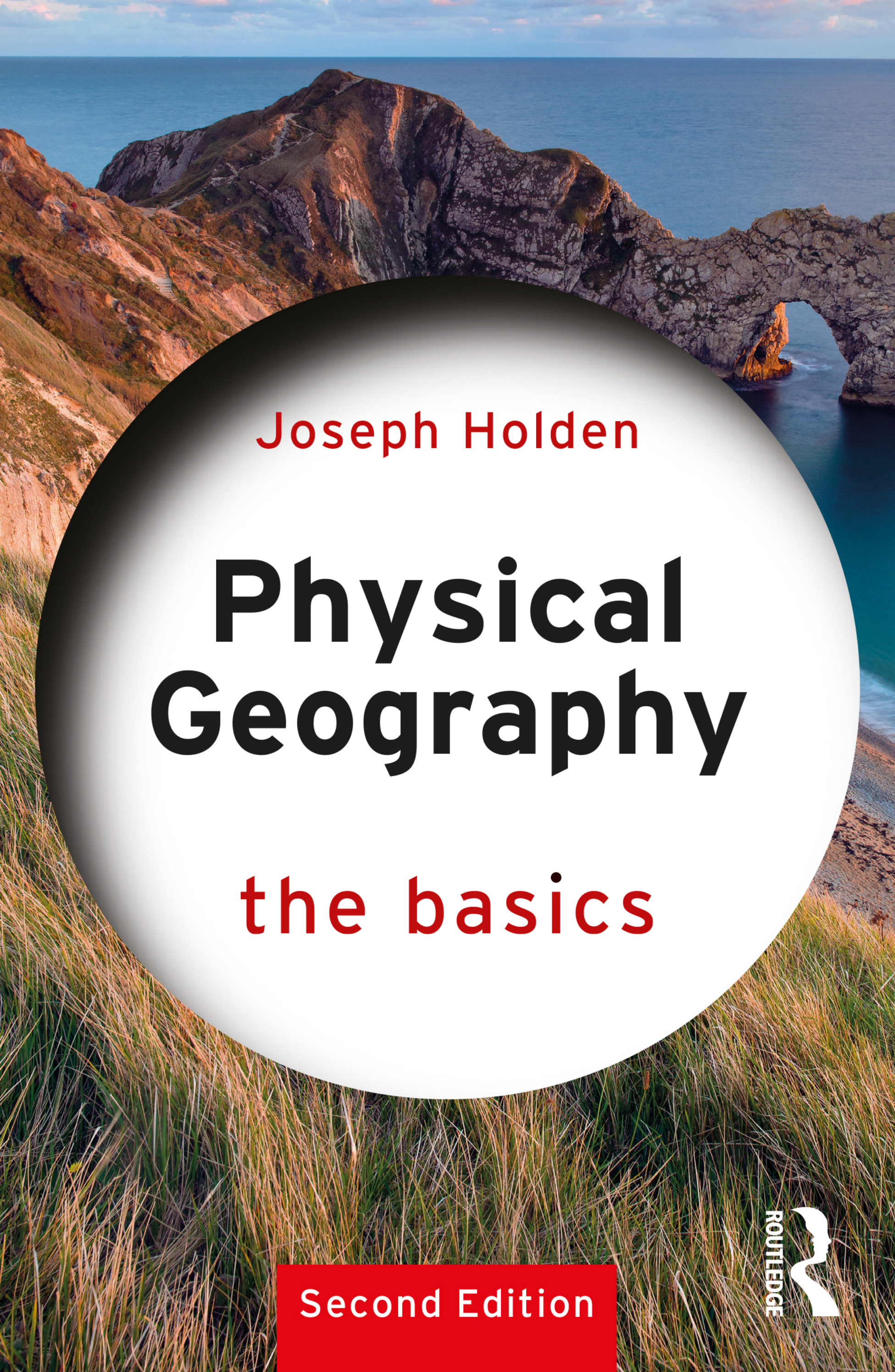

Joseph Holden

Physical Geography

the basics

Second Edition

ROUTLEDGE

PHYSICAL GEOGRAPHY

THE BASICS

This second edition, enhanced with more than 30 new figures, provides an up-to-date overview of physical geography suitable for all those with a personal or professional interest in environmental processes, climate change and understanding of the Earth's landforms and dynamics. The text provides explanations of processes, enabling the reader to understand the interconnected nature of the Earth's system, and has been updated to include new developments and case studies with insights from satellite observations and data analysis using artificial intelligence.

The book begins by outlining the nature of the Earth system, concepts around environmental thresholds and feedbacks, planetary boundaries for human survival, and humans as a dominant driver of environmental change. The second chapter examines features associated with plate tectonics, the role of weathering and erosion in shaping landscapes, and soil functions and management. Chapter 3 deals with the climate system, describing drivers of the major atmospheric and oceanic circulation systems, the natural greenhouse effect, and regional climate and weather experienced for different zones across the planet. The global carbon cycle and long-term climate change are considered in Chapter 4 before moving on to tackle the latest knowledge on contemporary and future climate change, its impacts, mitigation and adaptation. Chapter 5 facilitates key understanding of hydrology, river channel dynamics, water quality, coastal processes, glacier dynamics and cold region landforms while Chapter 6 deals with the distribution and patterns of life on Earth and of the underlying processes that result in these patterns. The book concludes with a brief overview of considerations for managing environmental change and hazards, and requirements for achieving the UN's Sustainable Development Goals.

This reader-friendly text brings together wide-ranging subject areas from across physical geography, covering the basics of the subject at a level suitable for those about to embark on a university degree or for those who just want to get a solid basic understanding of the physical environment around them. The book, which contains box features with examples and a glossary to aid understanding, acts as a primer for further study, or in itself can be used as a basic aid to understanding fundamental principles and processes associated with physical geography.

Joseph Holden holds the Chair of Physical Geography at the University of Leeds. He is Fellow of the Royal Geographical Society and Royal Meteorological Society and has won numerous awards. He has been Research Dean for the Faculty of Environment and is also Director of water@leeds, a major interdisciplinary water research centre.

The Basics

The Basics is a highly successful series of accessible guidebooks which provide an overview of the fundamental principles of a subject area in a jargon-free and undaunting format.

Intended for students approaching a subject for the first time, the books both introduce the essentials of a subject and provide an ideal springboard for further study. With over 50 titles spanning subjects from artificial intelligence (AI) to women's studies, *The Basics* are an ideal starting point for students seeking to understand a subject area.

Each text comes with recommendations for further study and gradually introduces the complexities and nuances within a subject.

MUSIC COGNITION: THE BASICS
HENKJAN HONING

PERFORMANCE STUDIES
ANDREEA S. MICU

AMERICAN STUDIES
ANDREW DIX

PHYSICAL GEOGRAPHY (SECOND EDITION)
JOSEPH HOLDEN

WORLD PREHISTORY
BRIAN M. FAGAN AND NADIA DURRANI

FRENCH REVOLUTION
DARIUS VON GÜTTNER

RESEARCH METHODS (THIRD EDITION)
NICHOLAS WALLIMAN

For a full list of titles in this series, please visit www.routledge.com/The-Basics/book-series/B

PHYSICAL GEOGRAPHY

THE BASICS

SECOND EDITION

Joseph Holden

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Second edition published 2022
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

First edition published by Routledge 2011

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Holden, Joseph, 1975- author.
Title: Physical geography : the basics / Joseph Holden.
Description: Second edition. | Milton Park, Abingdon, Oxon ; New York, NY : Routledge, 2022. |
Series: The basics | Includes bibliographical references and index.
Identifiers: LCCN 2021019953 (print) | LCCN 2021019954 (ebook) |
ISBN 9780367642716 (hardback) | ISBN 9780367644086 (paperback) |
ISBN 9781003124405 (ebook)
Subjects: LCSH: Physical geology.
Classification: LCC QE28.2 .H635 2022 (print) | LCC QE28.2 (ebook) |
DDC551--dc23
LC record available at https://lccn.loc.gov/2021019953
LC ebook record available at https://lccn.loc.gov/2021019954

ISBN: 978-0-367-64271-6 (hbk)
ISBN: 978-0-367-64408-6 (pbk)
ISBN: 978-1-003-12440-5 (ebk)

DOI: 10.4324/9781003124405

Typeset in Bembo
by Taylor & Francis Books

CONTENTS

5 Water and ice 106

FIGURES

TABLES

BOXES

ACKNOWLEDGEMENTS

Thousands of scientists have undertaken research to understand the way the world works, and in doing so have informed the content of this book. I hope that this book, which briefly summarises the outcomes of their dedication and endeavour, provides inspiration for the next generation to follow in their footsteps.

I would like to thank my family for their ongoing support, and in particular Eve Holden for her constant enthusiasm and my parents Henry Holden and Patricia Holden for their boundless positivity. Elizabeth Pettifer compiled much of the glossary and index and Alison Manson is thanked for preparing original versions of Figures 2.1, 2.7, 3.1, 3.2, 3.4, 3.5, 4.1, 4.2, 5.4, 5.6, 5.9 and 6.3. Justina Holden is thanked for updating or preparing Figures 2.2, 2.6, 3.4, 3.5, 4.1, 4.3, 5.2, 5.8, 6.1 and 6.3.

GLOSSARY

Ablation zone The part of the glacier where there is a net loss of mass annually.

Active layer The layer of ground above permafrost that is seasonally frozen and thawed.

Aeolian transport The movement of a substance by the wind.

Aerosols Microscopic particles contained within the atmosphere that interfere with the Earth's incoming energy from the Sun and the outgoing energy being emitted from the Earth's surface. Aerosols can result in either cooling or warming of the Earth's climate.

Albedo The proportion of incoming energy from the Sun that is reflected by a surface. Snow has a high albedo, whereas tarmac has a low albedo.

Anthropocene Geological epoch defining the period when humans were the dominant driver of global environmental change. The epoch has not yet been formally recognised by international geological committees.

Aquifer Rock that is porous enough to absorb and retain water and permeable enough to allow groundwater to pass through it freely.

Arête A narrow mountain ridge between two adjacent cirques.

Avulsion Sudden shift in the position of a river channel, abandoning the old course and taking a new one somewhere else on the floodplain.

Baseflow The stable portion of a river's discharge that is fed by groundwater.

Bases Substances that have a pH higher than 7 and release hydroxide ions (OH^-).

Berm The limit of the swash zone at the back of the beach, which exhibits a steep slope and a flat top.

Biodiversity The number and variety of different plant and animal species within an ecosystem.

Biogeography The study of spatial patterns and processes related to distributions of plants and animals and the interactions between living organisms and the environment.

Biomass The total mass of living organisms both below and above ground.

Carbon footprint The amount of carbon that an individual or an organisation uses over a given time period. It is often expressed in terms of the equivalent amount of carbon dioxide.

Cation exchange capacity A measure of a soil's ability to hold and release certain elements, which is dependent on the net negative charge of the clay minerals within it.

Chemosynthesis The use of energy from the oxidation of chemicals to manufacture carbohydrates for living organisms from water and carbon. The process uses chemical nutrients to produce energy for carbohydrate production rather than the Sun's energy, as used in photosynthesis.

Cirques Curved basins eroded into mountain sides by glaciers. They possess steep rock walls, a rock basin and sometimes a terminal moraine.

Cohesion The attraction between molecules in a substance that holds the substance together.

Col A hollow that forms in an arête from localised erosion, often providing a pass between mountain peaks.

Compound effects Where two or more environmental hazards combine and become enhanced to create a larger total impact.

Compounds Substances made up of two or more different elements, where the elements are chemically bound together.

Convection The molecular motion responsible for the transfer of energy, such as heat, through a fluid.

Convergent plate boundary Where two tectonic plates collide, creating major physical features, such as mountains, and hazards, such as earthquakes and volcanoes.

Coriolis effect The rotation of the Earth on its axis causes a moving object or fluid to be apparently deflected. Deflection occurs to the right in the northern hemisphere and to the left in the southern hemisphere, and is stronger at higher latitudes.

Corries See *cirques*.

Crag and tail features Formed as a glacier slides over a resistant rock mass ('crag'), depositing sediment on the downstream side ('tail') and creating a streamlined feature.

Critical zone That part of the Earth from the treetops down to the lowest level of groundwater – a zone where rock meets life and where rock, soil, water, air and living organisms interact.

Dew point The temperature at which a cooling air parcel becomes saturated with water vapour, at which water vapour condenses to form liquid water.

Divergent plate boundaries Where two tectonic plates move apart and create new crust, leading to features such as the ocean floor, a mid-ocean ridge and rift valleys.

Drumlins Streamlined mounds of till in the direction of ice flow, with a blunt upstream end and a tapered downstream end.

Dry adiabatic lapse rate Air is a poor conductor of heat, therefore the cooling or warming of an air parcel is considered to be adiabatic (self-contained), as there is little exchange of energy between the air parcel and the surrounding atmosphere. The dry adiabatic lapse rate is 9.8°C per kilometre and applies only to air parcels that have not reached the dew point, at which condensation occurs and the saturated adiabatic lapse rate is applied.

Ecological footprint An estimation of the amount of ecological resources used by an individual, company or nation.

Ecological niche Where there are no competitors for any of the resources required by an organism it can occupy the ideal conditions to which it is adapted within that ecological community.

Ecosystem services Services provided by ecosystems that support human life in some way, such as clean water, food and fibre, climate change mitigation and spiritual wellbeing. Evaluating these services helps to focus public attention on environmental issues that could result in the loss of these services.

El Niño Southern Oscillation A reduction in the upwelling of cold, deep water in the South American Pacific resulting from the reduced strength of trade winds over the equatorial Pacific Ocean, which in turn reduces the strength of westward-driven currents. This leads to unseasonal warm weather and disruption of pressure and precipitation systems in the southern hemisphere every 3–7 years.

Environmental lapse rate The normal rate of temperature change with altitude. Air temperature falls by approximately 6.4° C per kilometre, but this is subject to variation.

Erratics Large boulders that have been transported a long distance from their original source by glacial ice.

Eskers A long, narrow, wavy ridge containing fine material transported by R-channels at the glacier bed. Eskers can be 20 to 30 metres high and up to 500 kilometres in length.

Föhn wind The warm, dry wind that blows down the lee side (downwind) of a mountain, causing that area to be significantly drier. The dryness occurs because water held by an air mass will be released as it rises, cools and condenses over mountains to form precipitation.

Food chain The feeding links between species within an ecosystem indicating which species consume which other species.

Frost creep The downslope movement of the active layer. The soil expands perpendicular to the sloping surface when it freezes and settles vertically on thawing, causing an overall movement downslope.

Genetic modification The scientific alteration of DNA, where genes are deleted or added in order to change certain characteristics of a species.

Geomorphology The study of landforms and their characteristics so that their origin, development and history may be understood.

Glacials Long, cold phases during the Quaternary (last 2.4 million years), which saw the widespread advancement of glaciers and ice sheets.

Gleying The process by which a gley soil is produced. This happens when iron and manganese compounds within the soil are subject to stagnant, wet conditions and starved of oxygen so that the compounds are 'reduced'. The result is a blue-grey soil.

Groynes Long, narrow, artificial structures installed on beaches or river channels at right angles to the main sediment flow direction, which trap sediment and maintain the beach or banks.

Gyre A large, circular, rotating ocean current.

Hanging valleys A side valley to a main glacial valley that terminates at a higher elevation than the floor of the main valley due to glacial erosion in the main valley.

Herbivores An organism that consumes plant material only.

Holocene The last 11 700 years to the present day – an interglacial (warm period) of the Quaternary.

Humus Soil organic matter which is very resistant to decomposition.

Hydraulic conductivity A measure of the ability of water to flow through a substance, such as soil, rock or plant material, for a given hydraulic gradient.

Hydraulic gradient The difference in water level between two points across a horizontal distance measured in metres per metre.

Ice wedges A V-shaped wedge of ice that forms in areas with no insulating snow cover. When the ground becomes very cold it cracks to create frost crack polygons. The cracks become filled with water, which then freezes, creating a wedge of ice which can grow through time, expanding the crack.

Infiltration capacity The upper rate at which water can flow into the soil from the surface. This rate can change through time depending on how wet the soil is and the surface conditions.

Infiltration rate The time it takes for a unit depth of water added to a surface to enter the soil.

Infiltration-excess overland flow Where the rate of rainfall or irrigation water supply to the soil surface exceeds the infiltration capacity, leading to excess water flowing across the surface. This is also known as Hortonian overland flow.

Infrared radiation The energy that is released by all solids, liquids and gases as heat.

Interglacials Long, warm phases of the Quaternary (last 2.4 million years), in which glaciers and ice sheets receded and became limited to a few locations.

Intertropical convergence zone The region where the trade winds from the northern and southern hemispheres converge. Conditions are favourable for warm, moist, rising air, resulting in cloudiness and heavy rainfall.

Jet streams Narrow, high-speed winds caused by sharp temperature gradients, located within Rossby waves in the upper atmosphere. They can be thousands of kilometres long and hundreds of kilometres wide.

Karst A landscape shaped by the weathering of limestone rock, characterised by underground drainage tunnels, surface depressions and, sometimes, tall rock towers.

Kettle holes Depressions in the surface of a glacial sediment deposit caused by blocks of ice which become surrounded by sediment and, after melting, leave a hole in the sediment.

Keystone species A species which is connected to many levels of the food chain, which if lost could result in collapse of the food web and major loss of biodiversity.

Lapse rate The rate at which air temperature decreases with increasing altitude.

Laterisation Warm weather and plentiful rainfall result in fast weathering conditions and leaching of material within soils. Laterite soils are produced, which are often orange or red in colour. These soils often have little organic matter since decomposition and leaching removes the material quickly.

Leaching The removal of dissolved soil material vertically through the soil profile by surplus water.

Litter Soil organic matter consisting of decomposing plant and animal debris.

Longshore drift The transport of sediment along the coast by longshore currents that surge towards the beach at an oblique angle, followed by a backwash that transports sediment at right angles to the coast, resulting in a zig-zag movement of material along the coast.

Macropore flow The transfer of water through the soil between large pores greater than 0.1 millimetres in diameter.

Massive ice Isolated bands or lenses of ice several metres thick within soils.

Matrix flow The transfer of water through the soil between pores smaller than 0.1 millimetres in diameter.

Meander A curve in the path of a river channel.

Mineralisation The release of plant nutrients during decomposition of organic matter, which can then be used by other organisms.

Moraines Linear mounds of glacial till that have been transported by a glacier. They are classified according to the method of their deposition.

Natural greenhouse effect Atmospheric greenhouse gases, such as carbon dioxide and water vapour, absorb 90 per cent of the long-wave radiation emitted from Earth, resulting in an average global temperature approximately 35°C warmer than would be experienced without the natural levels of greenhouse gases present.

Neap tides When the Sun opposes the gravitational pull of the Moon on the Earth (i.e. the Earth is positioned between the Sun and the Moon) the tidal range is reduced, resulting in lower high tides and higher low tides. It usually occurs during the first and third quarters of the Moon.

Occluded front A faster, steeper cold front overtakes a slower warm front and lifts the mass of warm air upwards.

Organic farming Farming which prohibits or minimises the use of human-made fertilisers and pesticides, and relies on natural biological processes to maintain soil fertility and control pests.

Partial pressure The pressure of an individual gas within a mixture of gases.

Peds Naturally occurring clumps of soil particles.

Periglacial A region subject to cold temperatures with repeated freezing and thawing, often occurring in zones near ice sheets or glaciers.

Photosynthesis The process by which organisms such as plants and algae (autotrophs) create carbohydrates and release oxygen using light energy, carbon dioxide and water.

Pingos An ice-cored mound of earth that can reach 60 metres high and 500 metres in length, which are only found in periglacial areas. They are caused by the doming of the frozen ground beneath a former water body.

Plant functional types A way of classifying plant species based on their traits and association with specific environmental variables. Plant functional types are used by environmental models to predict how certain groups of species might respond to climate change.

Plate tectonics A theory that concerns the large-scale movements of the Earth's crust (lithosphere). The crust is split into several 'plates', categorised as either oceanic or continental, which are moved by convection within the liquid mantle upon which they float. The movement of the plates is responsible for mountain building, trenches and earthquakes.

Ploughing boulders Boulders found on periglacial slopes that move slowly downslope, pushing through the soil and leaving a trough behind them, which creates a bulge of sediment in front of the boulder. The movement is thought to occur due to the difference in thermal conditions beneath the boulder compared with its surroundings.

Polar front The boundary at which cold air from the polar cell and warm tropical air from the Ferrel cell meet, causing air to rise.

Precautionary principle A decision making approach which believes that lack of scientific evidence for warnings about future threats of serious damage should not be used as an excuse to avoid action in order to prevent damage from happening; action should be taken as early as possible.

Precipitation The condensation of water vapour to form water droplets in the atmosphere, which are then deposited on the Earth's surface in a liquid (e.g. rain) or solid form (e.g. snow, hail).

Pressure-melting point The melting point of solids, such as ice, is not constant but varies according to pressure. The melting point of ice becomes lower with increasing pressure, meaning that under thicker ice masses there is likely to be water.

Protalus ramparts A linear arrangement of coarse sediment at the base of a periglacial slope, which is created by frost shattering of slope material that then slides down over the snow pack and settles below it.

Quaternary The last 2.4 million years to the present day, characterised by the expansion and contraction of ice sheets in predictable cycles.

Refraction A process caused by a reduction in velocity as a wave enters shallow water, resulting in the wave front changing direction and 'bending' as it reaches the shore.

Regelation When ice meets an obstacle, such as a rock, pressure increases on the upstream side of the rock. This lowers the pressure melting point, resulting in melting of the ice on the upstream side of the rock, which then flows around the obstacle and refreezes on the downstream side due to the pressure melting point being higher. This allows ice to flow around obstacles.

Regolith The layer of soil overlying the bedrock that contains unconsolidated weathered parent material, which provides the raw material for soil development.

Riffles The accumulation of coarse sediment along shallow sections of river, which forms bar deposits across a river, and which tend to be spaced apart downstream between five and seven times the channel width.

River regime The variability of river flow over time, typically characterised over a year.

Roche mountonées Smaller versions of stoss-and-lee forms.

Rock glaciers A tongue-shaped body of rock and angular sediment that flows very slowly downslope like a glacier. Ice often occurs within pore spaces between the rock particles, which aids movement.

Rossby waves Large upper-atmosphere undulations that circumnavigate the globe and which disturb the belt of prevailing westerly winds associated with the Ferrel cell, causing a wavy distribution. They contain jet streams.

Safety factor (also known as factor of safety) The ratio of the forces resisting movement to the forces promoting movement of material downslope. If the value is below 1, movement will occur.

Salinisation The collection of soluble salts within a soil that can have a detrimental impact on soil fertility. It concerns salts of sodium, magnesium and calcium.

Salinity The concentration of salt dissolved in water.

Saltate A method of sediment transport whereby sediment grains are bounced along a bed surface.

Saturated adiabatic lapse rate See *dry adiabatic lapse rate* for explanation of ‘adiabatic’. The saturated adiabatic lapse rate is applied to air parcels that have reached the dew point, which causes water vapour within the air parcel to condense into liquid water. This process releases heat and warms the air parcel, meaning the saturated adiabatic lapse rate is less than the dry adiabatic lapse rate. The saturated adiabatic lapse rate varies according to the temperature and moisture content of the air.

Saturation-excess overland flow Where all of the pore spaces within the soil become filled with water, and therefore saturated, forcing the excess water to flow across the surface.

Segregated ice A lens of ice found just below the active layer, which grows because of the migration of water from around the lens to the freezing point.

Sesquioxides Contain three oxygen atoms and two atoms (radicals) from a different substance.

Shear stress A stress acting upon a particle in the same direction as the surface it is resting upon. In rivers the shear stress is the velocity of flowing water. When a sediment particle can be lifted from the river bed, then the critical flow velocity (critical shear stress) is reached.

Shoaling A gradual decrease in water depth that causes waves to become higher as they move towards the land.

Shore platforms A flat or gently sloping wave-cut platform found at the base of a rocky cliff cut into the rock.

Soil horizons Layers of soil within a soil profile, distinctive in terms of colour and texture, created by the leaching and deposition of soil materials that is caused by water moving vertically through the profile.

Solifluction The slow movement of soil material downslope, where a saturated thawed layer moves across a layer of permafrost under gravity.

Solute flux The total movement of dissolved material through a system as measured by mass (e.g. kilograms).

Speciation The evolution of a new species, whereby small changes in the characteristics of successive generations leads to a species that is different from the ancestor from which it originated.

Spits A narrow strip of sand that protrudes into the sea, usually curved in a seaward direction, with one end attached to the mainland. They occur where the shore direction changes, e.g. at the mouth of an estuary.

Spring tides A tide that occurs on or around the time of a new or full Moon, which is characterised by unusually high or low tides. It occurs when the Sun and Moon are in alignment, reinforcing the gravitational pull of the Moon on the Earth.

Stoss-and-lee forms Smoothed rock outcrops formed by basal sliding of a glacier, streamlined in the direction of glacier flow, with a plucked steep side facing downstream and a tapered end facing upstream.

Striations Small erosional features, with an appearance similar to scratches, caused by basal sliding of a glacier with embedded particles at the base of the ice, flowing over a rock surface and creating grooves.

Storm surge A rise in local sea level that occurs during storms where the wind forces water up against the coast. Storm surges mean that sea levels are much higher than can be accounted for by the tide alone and they can lead to coastal flooding.

Subduction zone Subduction can occur where two oceanic plates are forced together and one slides beneath the other, or where an oceanic plate is forced beneath the denser continental plate. The plate material is then melted by the mantle beneath the Earth's crust.

Sublimation The process by which a solid changes directly into a gas.

Succession The changes to the structure and make-up of an ecological community over time.

Surf zone The area where the water depth becomes too shallow for waves, causing them to break.

Tarn A small lake that occupies the basin of a cirque where a glacier was once located.

Temperature inversion Air temperature increases with altitude, usually where a layer of warmer air overrides a layer of cooler air. The reverse of the normal environmental lapse rate.

Thermohaline circulation Large-scale, deep-ocean circulation involving vertical and lateral movements of large parcels of water, driven by gradients of water density, which results from variations in water temperature and salinity.

Thermokarst A collection of irregularly spaced thaw lakes and depressions as a result of the melting of segregated ice.

Throughflow The movement of water flowing through the soil or bedrock.

Tidal bore The leading part of the incoming tide, which forms a wave under the special circumstances of a narrow bay and river channel, whereby the wave travels up river in the opposite direction to the river's normal flow.

Tidal current The flow of water that is produced by the rise and fall of the tides, which is most pronounced in river mouths, estuaries and where flow is squeezed through inlets.

Trade winds Strong winds that flow from east to west towards the equator, between 30° north and south. They are deflected to the west due to the Coriolis effect.

Tragedy of the commons A theory that explains how communal resources can be degraded due to the selfish nature of individuals who use more than their fair share. It is used as an analogy for the unsustainable use of finite ecological resources

Transpiration The evaporation of water through the pores of plant leaves, which is released into the atmosphere.

Trophic levels Groupings of organisms within a food chain (see entry for *food chain*). Primary producers (photosynthesising plants that form organic material from the Sun's energy, carbon dioxide and water) are at the lowest tropic level and these are fed upon by creatures at the second trophic level and so on.

Troposphere The lower layer of the atmosphere that extends between 6 and 15 kilometres in altitude above the Earth's surface.

Tsunami An energetic sea wave triggered by an earthquake, landslide or meteor impact in the ocean, which can become very large once it reaches shallow water and cause devastation in coastal zones.

Water table The upper limit of the saturated zone of the soil or rock.

Whaleback forms Smoothed rock outcrops formed by the sliding of a glacier, streamlined in the direction of glacier flow, with a smoothed steep side facing upstream and a tapered end facing downstream.

Yardangs Hills that have been smoothed and streamlined by erosion from dust carried by the wind.

1

INTRODUCTION

SCOPE OF THE BOOK

Physical geography involves the study of the flows of energy, water, nutrients and sediment that shape the Earth's landscapes and oceans, and the interactions between these flows and landscapes and the climate system, plants, animals and people. This book therefore brings together wide-ranging subject areas from across physical geography, covering the basics of the subject at a level suitable for those about to embark on a university degree or for those who just want to get a solid basic understanding of the physical environment around them. The book acts as a primer for further study, or in itself can be used as a basic aid to understanding fundamental principles and processes associated with physical geography. While there is coverage of different types of environments within the book, such as coasts, drylands, tundra, tropical forests and so on, primarily the book focuses on imparting a process understanding of how the Earth functions.

Material has been grouped into seven chapters. This first short chapter sets the scene and outlines some of the key concepts related to the interconnections between processes, ecosystems and landforms across the planet. The second chapter covers major processes that shape the Earth's physical appearance, including tectonic processes, weathering and erosion. It moves on to discuss soil processes: soils form as a result of weathering and biological action over thousands of years and are crucial for human survival. The Earth's climate system involves connections between the atmosphere, oceans, biosphere, cyrosphere (ice-dominated zones)

DOI: 10.4324/9781003124405-1

and the landscape. The major circulation systems are covered in Chapter 3, which also describes regional climate and weather for different zones across the planet. Chapter 4 then builds upon the understanding imparted by Chapters 2 and 3 by covering the global carbon cycle and outlining what we know about previous long-term climate change and contemporary climate change. The impacts of contemporary and future climate change are then covered, followed by an outline of some of the ways in which we might mitigate climate change and the need to adapt to climate change.

Climate change is impacting the water cycle, and changes in ice cover impacting the amount of the Sun's energy that is reflected back into space and the amount of sea-level rise. These topics are covered in Chapter 5, which focuses on water and ice, starting with an explanation of how water moves across and through landscapes, including groundwater and surface flow. Flood risk and river channel dynamics are covered before outlining how water quality varies naturally and with human action. Around 40 per cent of the world's population live at the coast, and coastal regions represent a key water-dominated landscape: the physical processes shaping coastal landforms are discussed along with how humans have altered coastal processes by changing water and sediment dynamics inland as well as along the coastline itself. Chapter 5 then moves on to focus on the world's ice masses, which shape landscapes and also act as sources of contemporary and future global sea-level rise related to climate change. After covering glacier dynamics, the chapter deals with glacial landforms before examining **periglacial** landforms (bold words in the main text are glossary terms that are mentioned for the first time) and processes associated with cold regions that do not have glaciers and ice sheets. Chapter 6 deals with **biogeography**, which examines the processes responsible for the spatial distribution of plants and animals. The chapter covers ecosystem processes, the key biomes on Earth and human impacts on ecology. Finally, Chapter 7 discusses the types of environmental solutions required, both now and in the future, including those to support delivery of the United Nations' Sustainable Development Goals.

Both remote sensing and techniques for dealing with big environmental datasets are key technological tools where geographers lead the way. Examples are provided throughout the book of the use of such techniques to aid in understanding the Earth's processes and associated hazards (e.g. volcanic eruptions, flooding), to determine the nature of environmental challenges and to support environmental solutions. Examples and boxed features to illustrate key points or case studies are provided within the book. The further reading lists at the end of each chapter can be explored to develop an understanding of the processes in more detail for those areas that interest you most.

THE INTERCONNECTED EARTH

The study of physical geography helps us understand how the processes related to the Earth's landforms, water systems, atmosphere and biosphere all interact (Figure 1.1), and how humans modify these interactions. Tectonic processes slowly build mountains and may temporarily change the composition of the atmosphere (e.g. gas or ash releases from eruptions may either warm or cool the Earth). They also change the circulation systems in the lower atmosphere (e.g. the growth of the Himalayas over millions of years has altered atmospheric circulation patterns, leading to an annual monsoon season). Weathering and erosion processes, including glaciers and ice sheets, help to denude mountains and also change the

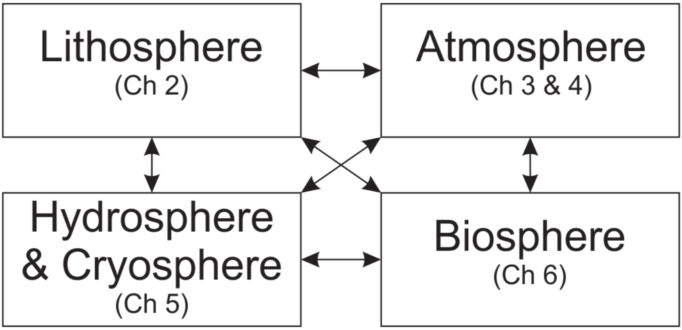

Figure 1.1 Key components of the Earth system and main chapters in this book that cover these components. Note that the linkages between the components are critical elements of physical geography and are emphasised throughout the book.

composition of the atmosphere (e.g. when rainfall dissolves rocks this takes carbon dioxide out of the atmosphere). Physical and chemical rock weathering is important for soil formation, but soil formation also requires biological activity. The soil conditions partly determine the suitability for plant growth while the plants often influence soil properties.

The **critical zone** is considered to be the part of the Earth that stretches from the treetops down through the soils and bedrock to the base of the groundwater zone. Scientists studying the critical zone work together across disciplines to examine all of the interactions within the zone, which are seen as critical for human life. There have been some critical zone observatories set up with infrastructure to monitor process dynamics and fluxes of energy, water, gases and nutrients, although long-term funding for such endeavours is difficult to maintain. There is also a critical zone research coordination network that brings together researchers to collaborate on studying the interconnected processes. The purpose is to be able to better forecast how the whole critical zone will respond to changes in climate, atmospheric chemistry, land management and water management.

There are countless *feedbacks* and connections at different scales operating across the Earth. For example, water flowing over a pebble in a river can cause small-scale erosion on the upstream side of the pebble, as water accelerates over the obstacle due to being squeezed past it. As the water moves more quickly it can pick up (entrain) fine particles from the river bed just upstream of the pebble. As the water slows down once it gets beyond the pebble, the fine sediment may be deposited. Therefore, the river bed may change shape around the pebble, with a deeper part upstream of the pebble and shallower part downstream. Once formed, the undulations in the river bed can grow larger as part of this positive feedback and start to influence water flow at a larger scale. The small dune feature may slow and accelerate water flow so that the dune enlarges to become a pool and **riffle** feature. The shallow riffle may cause water to start to flow around it, leading to bank erosion and the formation of a **meander** bend. A larger-scale example of feedback comes in the form of vast, deep-ocean circulation systems, which are influenced by water density differences. The ocean circulation systems are important for transferring vast amounts of energy around the planet and redistributing the heat from the tropics towards the poles. The water temperature and saltiness of the water influence its density, with more dense water sinking and less dense water rising. Evaporation from the ocean surface is one cause of increased water density because the salt concentrations will increase in the water left behind. However, where large-scale melting of ice sheets and glaciers occurs, large amounts of less dense fresh water may enter the ocean. If this happens at a location where normally the water becomes dense and sinks, then the vast turnover of deep-ocean circulation systems may be altered, changing how energy transfer occurs across the planet. A negative feedback may occur over hundreds or thousands of years if this shift in deep-ocean circulation results in the cooling of a region (less heat reaching that region from ocean transfers) that was previously experiencing warming and accelerated ice melt.

Understanding the interconnectedness of Earth system processes and feedbacks at different scales is crucial for our ability to manage environmental hazards, determine how human activities are impacting the Earth system and predict future changes to the Earth system. Through human action we have altered the nature and form of river beds and banks, and the river flow itself, and thus we have altered river channel dynamics. We have influenced ocean dynamics by changing the water quality of river water and by changing the composition of the atmosphere by burning fossil fuels and releasing other pollutants (gases are exchanged between the atmosphere and oceans). Land-based and aquatic ecosystems respond to changing atmospheric chemistry, but it is often difficult to understand all of the feedbacks because, at the same time, humans have made large changes to ecosystems on Earth through deforestation, urbanisation, agricultural land use and water impoundments, abstractions and diversions. Actions in one region (e.g. historic deforestation across Europe) can have global consequences (e.g. increased global carbon dioxide concentrations) or consequences in other regions (e.g. tropical deforestation leading to reduced rainfall a few hundred kilometres away). Therefore, projecting forward about the range of impacts of contemporary climate change and land and water modification is challenging, and requires thousands of researchers to pool their expertise from across different disciplines. However, such work is necessary so that we can plan and adapt to environmental change and deal with the changing nature of associated hazards (e.g. coastal flooding or wildfires).

ENVIRONMENTAL CHALLENGES

Major disasters, such as wildfires, earthquakes, volcanic eruptions, floods, droughts, landslides or tropical storms, can only be understood through integrated understanding of land–ocean–atmosphere–human interactions. Sometimes what seems to be a natural disaster (e.g. a major flood event) has been impacted by human activity (e.g. deforestation of hillsides in the upper parts of river basins), making the event worse. At the same time, humans have a creative ability to come up with solutions to major problems, or at least make things more resilient to disasters (e.g. buildings designed to allow them to stay standing during an earthquake). Solutions work best when the way the Earth's system works is fully understood.

However, with a growing world population the environmental challenges are ever increasing. Examples of these challenges include how we will feed an increasing population in a sustainable way without causing more pollution and greater release of greenhouse gases; how to deal with sea-level rise, given that a large proportion of the population live in coastal areas; how to reduce air pollution to avoid adverse effects on human health; how to slow the rate of species extinction, which is occurring at an unprecedented rate in human history; how to reduce soil degradation; how to ensure there is sufficient, clean freshwater for human use and for the natural environment; how to reduce the rate of ocean acidification; and how to build resilience to climate change impacts, such as increased storminess and droughts. Understanding climate change and its impacts is necessary so that we can adapt to changes and mitigate effects to minimise the damage to human life and infrastructure. Knowing about climate, soil, plant and water interactions is crucial to supplying food and clean water as the world's population grows from 7 billion to 9 billion over the next 30 years. Therefore, understanding the basics of physical geography should be a key area of knowledge and a fundamental area of understanding for the policy-making community.

A common concern is whether there are environmental *thresholds* beyond which a system will suddenly change its state, resulting in a sudden jump in the system or catastrophic event. Environmental systems are often thought to exist within certain ranges of tolerance, within which small-scale adjustments and feedbacks occur. However, it may be that a system can gradually change, perhaps via human action, towards a threshold (sometimes referred to as a 'tipping point'), at which point the system changes to a completely different state. From that point the system may not be able to return to its former state. Such an event may lead to huge impacts on humans in a particular region. For example, very slow accumulation of nutrients in a lagoon used by a local community for fishing can cause it to suddenly cross a threshold and, at that point, to change from a system with clear water and very diverse fish life, invertebrates and aquatic plants, to one which is murky, with lots of algae, low plant diversity and very few fish or invertebrates. At a global level, there may also be critical thresholds or boundaries beyond which the Earth's system may radically alter. These have been termed 'planetary boundaries' (Box 1.1).

BOX 1.1 PLANETARY BOUNDARIES

The planetary boundaries concept builds on a long-standing idea that there are 'safe operating limits' within which the Earth can sustain human life. If humans alter the system too much then the Earth may no longer be inhabitable. In 2009 a group of scientists, led by Johan Rockström, proposed that there were nine key systems essential for human life on Earth. They attempted to define the limits within which each of those systems might operate (Figure 1.2) so as to provide guidance for international governments. They acknowledged that there are large levels of uncertainty about these boundaries and the exact nature of the boundaries are not yet well understood. The nine limiting systems within the planetary boundaries framework are listed below.

Stratospheric ozone depletion: the ozone layer filters out ultraviolet radiation, and so if the ozone layer decreases it may allow more ultraviolet radiation to reach the Earth's surface, damaging ecosystems and causing human skin cancer. The hole in the Antarctic ozone layer detected in the 1980s showed that CFCs being released

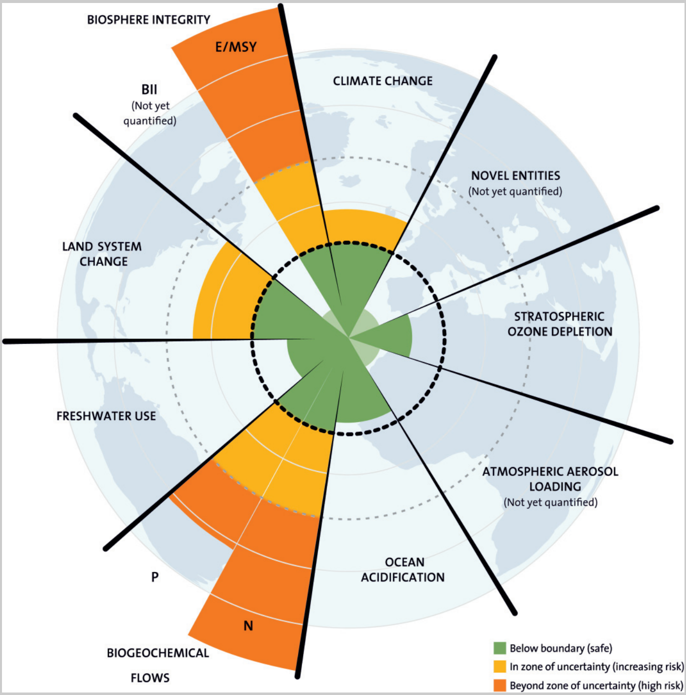

Figure 1.2 The nine proposed planetary boundaries for human life illustrating the potential for the current state of the Earth system to be below or above the boundary. E/MSY = extinctions per million species years (number of species out of every million species that become extinct per year). BII = biodiversity intactness index. P = phosphorus. N = nitrogen. (Source: J. Lokrantz/Azote based on Steffen et al. (2015) Planetary boundaries: guiding human development on a changing planet. *Science*, 347: 1259855.)

by humans were causing catastrophic damage, but quick action to stop further release of these pollutants has allowed the system to stay within the planetary boundary.

Loss of biosphere integrity (biodiversity loss and extinctions): current high rates of ecosystem damage and species extinction are unprecedented in human timescales.

Chemical pollution and the release of novel entities: emissions of toxic substances, including heavy metals and radioactive particles, are key risks to human life. At present, the planetary boundaries group have not quantified a single chemical pollution boundary for human life, but they note that crossing Earth system thresholds is considered sufficiently well-defined that it should be included as a priority for precautionary action and for further research.

Climate change: the planetary boundaries team suggest we have already crossed the threshold for high carbon dioxide concentrations in the atmosphere, which will eventually lead to catastrophic change, including major sea-level rise and loss of summer polar sea ice. The team suggest that a major unknown is how long the Earth's system can remain over this boundary before large, irreversible changes become unavoidable.

Ocean acidification: increased acidity in the oceans (by 30 per cent compared to pre-industrial times), caused by too much carbon dioxide in the atmosphere, makes it difficult for organisms such as corals and some shellfish and plankton species to survive. Loss of these species would change ocean ecosystems and could potentially lead to drastic reductions in fish stocks.

Freshwater consumption and the global hydrological cycle: climate change influences the water cycle, but direct management of freshwater by humans is now the dominant issue, changing global-scale river flows and, through vegetation change, altering the release of water into the atmosphere. A planetary freshwater boundary related to water use, water scarcity and environmental flow requirements has been proposed to maintain the overall resilience of the Earth's system.

Land system change: as a key driving force behind major reductions in biodiversity, changes in water flows and biogeochemical cycling of carbon, nitrogen, phosphorus and other important elements, the planetary boundary team recommend that the land system change boundary needs to reflect the quantity, function, quality and spatial distribution of a land system type. As such, they have chosen to focus on forest cover.

Nitrogen and phosphorus flows to the biosphere and oceans: humans have drastically altered the cycling of nitrogen and phosphorus through industrial and agricultural processes. The redistribution and change in form of nitrogen and phosphorus are causing major transformations to terrestrial, aquatic and atmospheric functions.

Atmospheric aerosol loading: **aerosols** strongly impact on Earth's climate system by changing how much solar radiation is reflected or absorbed in the atmosphere. Aerosols also have many adverse effects on living organisms, including human health. However, the relationships between aerosols and climate are complex and so the team have not yet defined the planetary boundary, although they have included it on the list as it is an important area for further research.

The planetary boundaries framework is based on the notion that human actions have become the main driver of global environmental change, so much so that it has been proposed by some that we are in a new geological epoch known as the '**Anthropocene**'. There is still debate about the precise starting point of the Anthropocene, when humans became the clear dominant driver of environmental change. Suggestions for the onset date have included the time of the earliest detectable human impacts, such as the start of organised agriculture several thousand years ago, the Industrial Revolution in the 1800s, when humans began to have a strong influence on atmospheric carbon concentrations, and 1945, when atomic bomb testing and deployment commenced. Two dates were proposed by Simon Lewis and Mark Maslin (2015), which they claimed met many of the criteria for defining the start of an epoch, including a sudden change in the geological strata record that could be identified globally. These were 1610 and 1964. The former was proposed because there was a clear global signal in carbon dioxide concentration changes in the atmosphere caused by humans. The 1610 signal was a low point in atmospheric carbon dioxide concentrations (detected in bubbles from ice cores at the time the ice was formed) caused by a massive decline in farming in the Americas due to the death of 50 million indigenous people from diseases such as small pox and influenza, which were introduced by European colonialists and to which the local population had no immunity, and from persecution by European colonialists. Major transfers of crops and animals between the continents also began at this time, denoting the start of permanent ecosystem changes on a global scale. For example, maize (a Latin American species) pollen appears for the first time in the European marine stratigraphic record in sediment layers from this period. The year 1964 was proposed as an alternative starting point for a formal definition of the Anthropocene as it coincides with the peak in atomic bomb carbon deposited in the global stratigraphic record.

SUMMARY

- Physical geography involves understanding spatial patterns (e.g. of landforms, weather, ecosystems) around the Earth through the study of flows of energy and matter across and between the Earth's atmosphere, biosphere, lithosphere and hydrocryosphere.
- There are many feedbacks and connections at different scales operating across the Earth. Understanding the interconnectedness of Earth system processes at different scales is crucial for our ability to manage environmental hazards, determine how human activities are impacting the Earth system and predict future changes to the Earth system.
- With a growing world population and increased use of natural resources, environmental challenges facing humans are ever increasing.
- A common concern is understanding whether there are environmental thresholds at different scales beyond which ecosystems or components of the Earth system will suddenly change their state, causing major problems.
- The Anthropocene epoch defines the period when humans became the clear dominant driver of environmental change. The years 1610 and 1964 are two key proposed dates for the onset of the epoch.

FURTHER READING

Holden, J. (2017) Approaching physical geography. In Holden, J. (ed) *An introduction to physical geography and the environment* (4th edition) (pp. 3–26). Harlow: Pearson Education.

A chapter that outlines more about the nature of physical geography and methodological approaches to the subject.

Lewis, S.L. and Maslin, M.A. (2015) Defining the Anthropocene. *Nature, 519*: 171–180.

A useful article that reviews the concept of the Anthropocene and the evidence for different starting points.

Steffen, W., Sanderson, R.A., Tyson, P.D., Jäger, J., Matson, P.A., Moore III, B., Oldfield, F., Richardson, K., Schellnhuber, H.-J., Turner, B.L., Wasson, R.J. (2004) *Global change and the Earth system: a planet under pressure*. London: Springer.

This is a case-study rich book illustrating interconnectedness, feedbacks and thresholds. The book is free to download.

2

TECTONICS, WEATHERING, EROSION AND SOILS

On a large scale, the landscape and oceans change slowly through time, mainly through tectonic processes, weathering and erosion. It may take millions of years for mountains to form. Mount Everest, the world's highest peak, is around 50 to 60 million years old and is currently growing at around 4 millimetres per year. However, these slow processes can also lead to sudden change, which can present major hazards to humans. For example, movements in the Earth's plates can result in catastrophic earthquakes and volcanic eruptions. Slow weathering processes can weaken rocks, potentially leading to hazardous landslides. Humans also modify the environment, which can enhance hazards resulting from weathering and erosion. For example, vibrations from high-speed trains might cause local slope failures to occur. Removal of native vegetation for agricultural activity may increase soil erosion that, in turn, pollutes rivers or leads to dust storms.

In this chapter we start by covering tectonic processes, which build and destroy landscapes and oceans on a large scale. The chapter then focuses on weathering and erosion processes. While weathering, erosion and subsequent deposition of eroded sediment can also build landscapes (e.g. sand dunes at the coast or in deserts), the overall net effect on the continents is that tectonic processes build landscapes while weathering and erosion can ultimately destroy landscapes. The environment that we see today represents one point in the overall evolution of the landscape, which changes through time. The chapter then discusses soils and soil processes, which are vital to our lives as they provide the growth medium for a large proportion of our food globally. Soils are a product of

DOI: 10.4324/9781003124405-2

weathering and biological inputs, and they slowly accumulate through time. However, human action can cause rapid and widespread degradation of soils, which threatens food security. Therefore, careful soil management is required to ensure sustainable food supplies, good water quality and reduced loss of soil carbon to the atmosphere.

TECTONICS: CONTINENTS AND OCEANS

Looking at a map of the world it seems like the shapes of some of the continents should fit together (e.g. South America and Africa). This pattern was noticed by explorers centuries ago. During the nineteenth and early twentieth centuries, scientists such as Alfred Wegener examined evidence from fossils that suggested that all of the continents were once joined together in a single land mass, called Pangea, which slowly broke and drifted apart. However, it was only through surveys of the ocean floor, as part of naval submarine and nuclear research during the 1950s and 1960s, that data were produced that showed how the continents really did move slowly around the Earth and that the oceans' floors spread from their centres.

Detailed maps of the ocean's floors showed that there are large mountain ranges running through the centre of the world's major oceans. It was also found that the deepest parts of the oceans are located very close to the edge of the ocean rather than in the middle (Figure 2.1). Ocean floors were found to have 'magnetic stripes'. It is known that the Earth's magnetic field reverses every few hundred thousand years, and the direction of the poles is recorded at the time when volcanic lava forms and cools. Alternating north or south-facing magnetic stripes occur right across the oceans and are orientated parallel to the mid-ocean mountain ridges (Figure 2.1), showing that ocean floors formed at the mid-ocean and then moved slowly away on either side of the mid-ocean towards the continents. This provided the first real evidence that large masses of rock on Earth can slowly drift.

The Earth is roughly spherical, although it is slightly flatter at the poles, meaning that the journey is 42 kilometres shorter if you travel around the planet via the poles compared to traveling around the equator. The Earth has an inner core that is 1 200

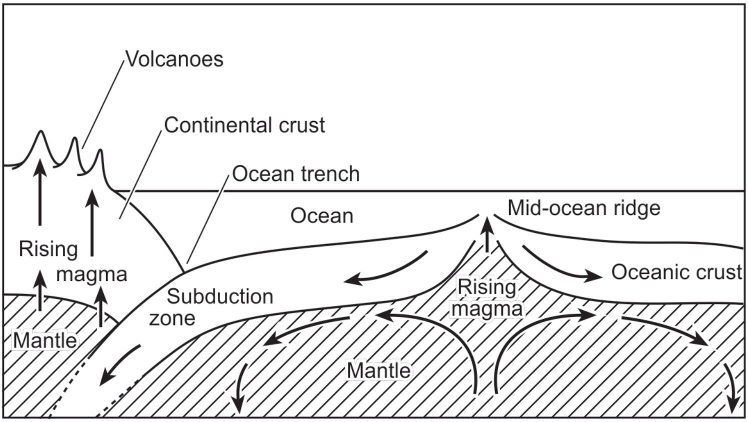

Figure 2.1 Mid-ocean ridge, sea floor spreading and ocean trench subduction zones.

kilometres thick, made of solid, hot (3 000°C) iron. The inner core is surrounded by a 2 300-kilometre-thick layer of liquid, iron-rich material, which forms the outer core. The next layer, as we move out from the centre of the Earth, is called the mantle (2 900 kilometres thick). The outer part of the mantle (180 kilometres), along with the overlying crust, is called the lithosphere and this is rigid and floats on the more mobile asthenosphere. Large continents form a cold and rigid crust, mainly made from granite rock, around 35 to 70 kilometres thick. Continental rock tends to be very old, often between 2 to 4 billion years. However, the rock on the ocean floors is made from volcanic basalt and forms a thinner, 6 to 10 kilometres thick, crust. This rock is much younger, with the oldest ocean floor being around 208 million years old.

Within the mantle, there are massive, hot currents of liquid kept molten by radioactivity within the Earth. These currents form circulation cells within the mantle (Figure 2.1). When the rising currents reach the stiff lithosphere above, they drag it along, causing the movement of the tectonic plates. In places, the hot, rising mantle material can force its way through the crust of the sea floor. Here, underwater volcanoes develop, forming a deep-sea mountain range (mid-ocean ridge). The lava from these underwater volcanoes forms new crust as it cools. As the descending parts of the convection cell in the mantle separate, they can slowly (1–10 centimetres per year) drag overlying ocean crust with them. The great, deep trenches around the edges of the oceans are zones where old crust is forced down and melts into the Earth's mantle (Figure 2.1). The ocean crust is, therefore, continuously being recycled, with new crust in the centre of the oceans and the thickest and oldest rocks on the ocean floor found nearest to the continents.

There are several plates moving across the Earth's surface (Figure 2.2). The continents are quite passive features of these moving plates since they just ride on top of them and, unlike ocean floors, they are not consumed into the mantle. However, when continental rock on two plates collides, material can be forced upwards, forming huge mountain chains. The boundaries between plates are therefore associated with different environmental features depending on whether the plates are moving apart at that boundary, coming together or sliding past each other. Earthquakes occur most at the boundaries of the plates, which is why we find that there are some locations that are much more prone to earthquakes than others. While plates slowly move you can imagine them having rough or bumpy sticking points. Over time, an enormous force builds up and eventually the plates move in a jolt, which is experienced as an earthquake.

Movements at the boundary between two plates can explain the nature of landforms found in these areas. Where plates are moving apart there are **divergent plate boundaries** (e.g. at the mid-ocean ridge) where new crust is formed. The lava formed at mid-ocean ridges is hot and very runny, forming gently sloping shield volcanoes. Volcanic eruptions with this type of lava (e.g. on Iceland, which straddles a mid-ocean ridge) tend not to be explosive because gas bubbles can easily escape through the runny liquid, although occasionally large gas bubbles do emerge, creating a scene with runny lava flying into the air. Eruptions may be in the form of walls of molten lava issuing from a linear crack in the Earth. Most divergent boundaries are in the mid-ocean, but there are some within continents. The Syrian–African Rift Valley is a good example of a divergent plate boundary on land. As the valley has continued to deepen, it is now below sea level and some of it has filled with water (e.g. the Dead Sea is 339 metres below sea level).

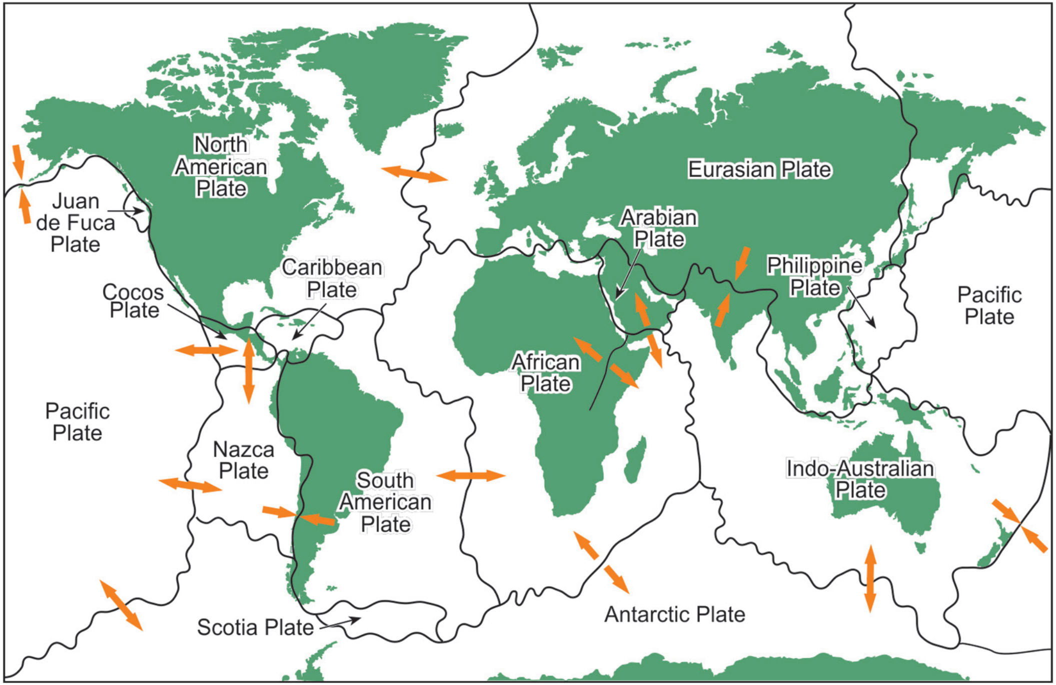

Figure 2.2 The main tectonic plates, with orange arrows showing directions of plate movements.

Transform faults occur where plates slide past one another (e.g. San Andreas Fault, California). Here there is often little creation or destruction of lithosphere and there are few volcanoes at transform boundaries. However, these boundaries can be associated with frequent major and destructive earthquakes. The rates of movement can be from a few centimetres in a small earthquake to two metres in a large event.

When two plates move towards one another, at a **convergent plate boundary**, major physical features are formed. If one of the plates slides beneath the other, a **subduction zone** is formed. This happens where two ocean crusts collide, or where denser ocean crust meets less dense continental crust and ocean crust is then forced under the continental crust and becomes part of the mantle. This is why ocean crust is relatively young by geological timescales. Often this process also creates mountain belts as the crust thickens at the subduction zone. For example, the Nazca plate collides with the South American plate (Figure 2.2) and is subducted below it, creating the Andes Mountains and many volcanoes.

Volcanoes produced around subduction zones can be very explosive and destructive. The oceanic crust is heated as it is carried down into the mantle. Water and other materials that are carried down with the plate are released, producing a mixture that rises to the surface. If the uppermost plate is oceanic, then basaltic volcanoes are produced, which form into arcs of islands. When an oceanic plate collides directly with continental crust, the oceanic plate moves under the continent. Water becomes trapped and causes the basaltic rock to melt under pressure. Rising magma starts to melt the continental crust of the overlying plate. This magma is very sticky and can result in a destructive volcanic explosion destroying large areas and killing many people. Examples of such volcanoes include Krakatoa (Indonesia), Vesuvius (Italy), Fujiyama (Japan) and Mount Saint Helens (USA). Sticky, slow-moving lava builds within these steep-sided volcanoes and once the lava ceases to flow, it cools, producing a plug allowing a considerable pressure to build up within the volcano ready for the next eruption.

There are around 1 500 potentially active volcanoes above the land or ocean surface (but tens of thousands on the ocean floor). On average, around 50 surface volcanoes erupt per year, although we tend only to hear about those that cause major destruction or disruption to travel. Recent work has developed satellite image analysis techniques to detect changes that may indicate an impending eruption (Box 2.1). Volcanoes form at the end of a central tube or vent that rises from the upper mantle. There is often a crater, which is a surface depression at the top of the volcano. Magma within the volcano slowly rises, building up pressure until conditions are right for eruption. Often the heat from the magma can also boil water in the ground, resulting in hot springs and geysers. Some of these form tourist attractions, such as Beppu in Japan, which has 2 500 public natural baths, or Wai-O-Tapu in New Zealand, which has colourful hot pools and geysers that jet boiling water 20 metres into the air.

There are some volcanoes that are not located at plate boundaries. These are at hot spots. They are probably located at the top of hot plumes within the mantle. There are over 40 hot spots and many of these have formed in the middle of plates. For example, at Yellowstone National Park, USA, there is a hot spot that is responsible for amazing geysers, mud pots and other volcanic features. There is evidence that there was a massive volcanic eruption there 600 000 years ago, depositing ash 12 metres thick up to 1 200 kilometres away. Where hot spots occur under the ocean, and where there is also a volcano, as the plate moves over the hot spot a series of volcanic islands or seamounts (underwater mountains that have not reached the water surface) are formed. The best example of such a system is the Hawaiian island chain. The island of Hawaii is furthest to the east and is the site of the most active volcano on Earth at present. In fact, the volcano, Mauna Loa, is the tallest structure on Earth when measured from its base on the ocean floor, at 10 kilometres tall. Mauna Loa and the whole island of Hawaii has only taken 1 million years to form, which is very short on geological timescales. Then there are a series of islands, Maui, Molokai, Oahu and Kauai, in a line stretching more than 500 kilometres to the west. Each island is made up of an extinct volcano, with the volcanoes becoming progressively older as you travel west. Further west, the islands are so old and eroded that they disappear below the ocean water surface to become seamounts. This also happens because the westerly moving ocean floor becomes deeper towards the ocean trench.

BOX 2.1 FORECASTING VOLCANIC ERUPTIONS FROM SPACE

Around 800 million people live within 100 kilometres of a volcano, and volcanic eruptions have often caused large losses of life. Sometimes volcanoes can be dormant for hundreds of years before erupting, and many lack ground-based monitoring. To protect humans, it is important to provide warning of an impending eruption so that necessary precautions can be taken, such as evacuation of the surrounding area. Satellite sensors can detect changes in ground movement, ground temperature and the density of the rock

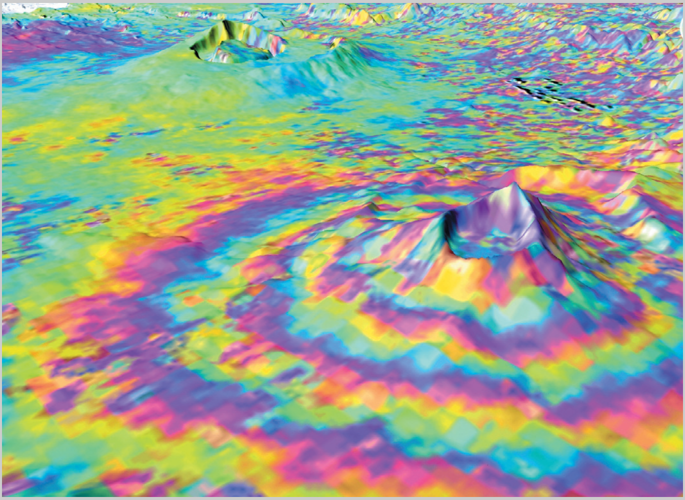

Figure 2.3 Uplift of the ground surface around the Longonot volcano, Kenya, in the foreground is indicated by the rainbow pattern in the interferogram produced by measured differences in the Earth's surface compared to the position of the satellite. No uplift was measured for the Suswa volcano in the background for the same period. The number of full cycles through the coloured bands indicates the size of the uplift, here equating to around 9 cm on Longonot before it erupted.

(Source: European Space Agency, www.esa.int/ESA_Multimedia/Images/2010/06/Volcanic_uplift. Creative Commons CC BY-SA 3.0 IGO: https://creativecommons.org/licenses/by-sa/3.0/igo.)

(by measuring gravitational fields), all of which may be important indicators of an impending eruption.

Before an eruption, magma tends to rise into chambers within the volcano, causing the ground surface to bulge. Using Sentinel-1 satellite radar data from the European Space Agency, it is possible to determine whether the ground surface position is moving, even by just a few millimetres. The satellite passes over each part of the Earth every few days. Radar waves are sent down to the ground surface and the satellite measures the energy reflected back. Small changes in the ground position between two satellite survey dates cause changes in the radar signal phase (because the distance between the ground surface and satellite has changed) and these can be assessed on coloured images known as interferograms, as shown in Figure 2.3. A lot of research by organisations such as the UK's Centre for Observation and Modelling of Earthquakes, Volcanoes and Tectonics, has been undertaken on computing techniques that can automatically filter and analyse radar data to detect ground movement.

For useful information and videos about satellite detection of change related to tectonic processes please see https://comet.nerc.ac.uk.

Convergent boundaries compress rocks and deform them. This compression causes the rock to fold and rumple as if it were a piece of cloth being pushed together from both ends, often producing mountain ranges that look like ripples when viewed from a plane high above. The most intense mountain building occurs when two continents collide. Colliding continental crust is not subducted and so it thickens and forces land to rise, thereby creating a large mountain belt such as the Alps, where Italy has moved north into Europe, or the Himalayas, where India has collided into Asia over the past 50 million years. The Himalayan continental collision zone seems to have shortened the length of continental crust by 1 000 kilometres. Therefore, here the crust is thickened and compressed and, as a result, the rocks are folded and deformed, crumpled and faulted. The thickened crust sticks up like an iceberg floating on top of the mantle with a deep root. All ten of the Earth's highest peaks are found in the Himalayas. As the high mountains are eroded and mass removed, then the floating root at the bottom moves up and exposes more rocks and minerals, which have been altered by high temperatures and pressures.

WEATHERING AND EROSION

Weathering is the physical breakdown of rock, whereas erosion is the transportation of the weathered material. Mountains that are built by tectonic processes are worn down by weathering and erosion. The sediments produced are moved around, often over vast distances, by water, ice or wind, and can be reincorporated into rock formation again over long timescales. Thus, the Earth's surface is in a constant state of change, albeit often too slow to be noticed within a human lifetime. The rates of change vary with rock type, climate, slope conditions, ice cover and vegetation cover.

TYPES OF ROCK

There are three main types of rock found around the Earth's surface: igneous, sedimentary and metamorphic rocks. Temperature at the time of formation, the mix of minerals present and pressure all interact to create varieties of these main rock types. Igneous rocks are formed when molten lava cools and hardens. If the molten rock is from a volcano, then the subsequent cooled and hardened basalt has small crystals. If the rock is able to cool slowly then larger crystals may grow, producing a coarse-grained rock such as granite.

Sedimentary rocks are produced by weathering of rocks followed by subsequent erosion and deposition of material. The deposited sediment can accumulate and eventually build up before being compacted and hardened over millions of years by the weight and pressure of sediments above and internal chemical changes. Rocks such as sandstone, siltstone or shales are good examples of sedimentary rocks. These rocks often contain a record of the physical conditions present when the rocks were deposited, including fossils. In fact, some rocks such as chalk or coal are almost entirely made from the remains of animals and plants.

Metamorphic rocks form through the partial melting and reforming of existing sedimentary or igneous rocks often under high pressure. Limestone and shale change to marble and slate when metamorphosed, for example. These rocks tend be more resistant to weathering as they are harder than other rock types.

The rock cycle means that all rock types can convert into other types. All rocks can be melted and cooled to form igneous rock. All rock types can be weathered and eroded to form the layers of sediment that can eventually become sedimentary rocks. Under pressure and heat, igneous and sedimentary rocks can transform to metamorphic rocks.

WEATHERING

Weathering is the breakdown of rocks by physical and chemical processes. These processes often work together, while biological processes can support weathering by influencing physical and chemical action. Some types of rock are more difficult to break down than other types of rock and so weathering can result in interesting landforms, such as headlands or areas of protruding rock in an otherwise flat landscape (e.g. Uluru, Northern Territory, Australia), where one type of rock wears down more quickly than another in the vicinity.

PHYSICAL WEATHERING

Physical weathering transforms rock by breaking it into smaller fragments through mechanisms such as freeze-thaw, salt weathering and thermal cracking. Physical biological action through roots, which force openings in rock, can also be important. Freeze-thaw is the process by which water freezes in small cracks and expands by 9 per cent as it does so. This then forces cracks to open further, eventually splitting the rock. The process is more active where temperature frequently fluctuates above and below 0°C.

Salt weathering occurs where salts within the environment form crystals in small cracks under desert conditions. The relatively bare rock surfaces in deserts, along with large diurnal temperature ranges and the lack of rainfall, can lead to salts becoming concentrated in surface locations and then cracking or flaking off the surface rock. Crystallisation occurs when temperature increases lead to the growth of salt crystals. Moisture inputs then cause the salt volume to increase, and thermal expansion also occurs to the salts upon warming. This type of weathering is more common where coastal fogs can bring sea salts into desert areas, such as in the Namib desert in Namibia.

Thermal weathering may also be important in deserts where daily temperature ranges may be high. This leads to the regular expansion and contraction of rock as it warms and cools. Different minerals expand by different amounts when warmed. This differential expansion creates internal stresses that weaken rocks and loosen particles. Rocks can sometimes crack if you light a fire around them due to these expansion stresses.

Tors (e.g. Figure 2.4) are towers of jointed, often broken, slabs of granite (and sometimes other rock types). The blocks of stone balanced on top of each other can falsely give the impression that humans made the structures. A range of weathering processes are thought to lead to their formation, but tors are the more resistant remnants of the surrounding landscape that has since been lost to weathering and erosion. Tors form over hundreds of thousands of years. In some regions, freeze-thaw has been a key factor in tor formation, whereas in other

Figure 2.4 Kit Mikayi, a 40-metre high tor, 30 kilometres east of Kisumu, Kenya.

(Source: Valerius Tygart, CC BY-SA 3.0: https://creativecommons.org/licenses/by-sa/3.0/deed.en.)

areas (e.g. Figure 2.4) salt weathering has been more important. The weathering processes most successfully work on the slightly weaker joints that were inbuilt into the rock structure since formation. These lines of weakness are preferentially removed over long time periods, leaving blocks balancing upon one another. Some of these amazing formations, such as that shown in Figure 2.4, have sometimes been of human cultural and religious significance.

CHEMICAL WEATHERING

Water acts as a solvent to dissolve rock. Rocks are made up of **bases** (calcium, magnesium, sodium and potassium), silica and **sesquioxides** (mainly with aluminium). Silica is at least ten times less soluble than bases, but is at least ten times more soluble than the sesquioxides. Therefore, chemical weathering reduces the proportion of bases the most, followed by the proportion of silica. As material is exposed close to the surface, the atmosphere is able to assist weathering of minerals. The gases of the atmosphere, such as oxygen, water vapour and carbon dioxide, aid weathering (e.g. iron and oxygen can produce iron oxide (rust)). Small amounts of carbon dioxide from the atmosphere dissolved in rainwater make a weak carbonic acid, which acts to weather rock. More intense rainfall combined with warmer temperatures and higher carbon dioxide concentrations may increase chemical weathering rates under climate change. However, when the carbonic acid reacts with the rock in the weathering process it produces other dissolved chemicals which then get transported away in solution. This can result in loss of the carbon dioxide from the atmosphere into oceans via river channels, and could potentially act as a negative feedback on climate change, but on very long timescales.

Chemical weathering can be assisted by biological action. For example, vegetation roots release organic acids into the soil, which help the roots to extract nutrients needed by the plant. However, these acids also help dissolve soil and rocks. Many of the small animals in the soil (e.g. earthworms) pass soil material through their bodies, altering it both biochemically and mechanically as they extract nutrients from it.

Karst refers to landscapes that are strongly shaped by dissolving very soluble rocks, such as limestone, gypsum and dolomite. Around

15 per cent of the Earth's surface has karst landforms, which are home to 1.2 billion people. Rapid karst development in wet tropical areas can result in amazing rock towers, such as those shown in Figure 2.5a, where most of the surrounding landscape has been dissolved, leaving tall narrow peaks between. In colder regions, glaciation can scour away soil cover, and once the ice has retreated this can leave behind exposed rock, which can then undergo solutional weathering preferentially along the joints of weakness. Once the joint weathering is initiated water can accumulate more readily along the depressions and accelerate weathering even further, resulting in a landscape with a series of small slabs of rock, known as clints, separated by deep weathered grooves, known as grykes. Such features are often known as 'limestone pavement' (Figure 2.5b). Often the subsurface of karst areas also undergoes significant weathering and erosion, creating large voids and surface collapse features

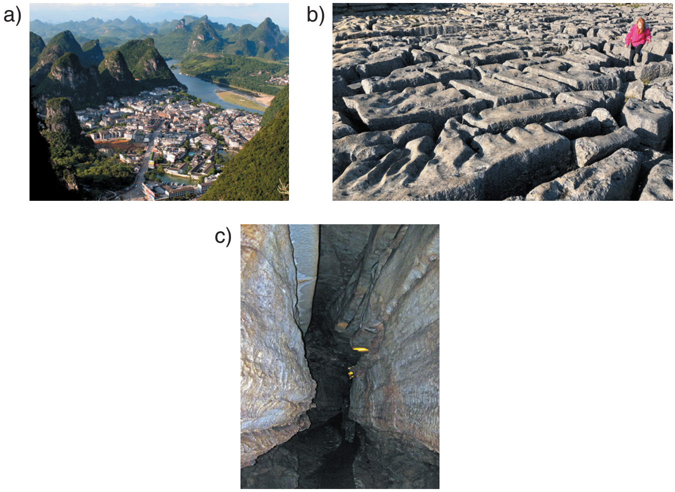

Figure 2.5 Karst landscape features: a) towers of karst around the town of Yangshuo, China; b) limestone pavement at Malham, England; c) underground passage created by a river flowing through weathered limestone in Mammoth Cave, Kentucky, the world's longest mapped cave system.

(Source: a: Ericbolz; c: James St. John, CC BY 2.0: https://creativecommons.org/licenses/by/2.0.)

such as sink holes and depressions. Water flow can be concentrated along weaker joints, which can cause more rapid dissolution of the joints, forming underground passages. The passages can evolve into large cave systems (Figure 2.5c). Streams can run underground along these passages for parts of their course in karst landscapes, sometimes for tens of kilometres, emerging at springs. Hot desert areas have the slowest karst development due to a lack of water for solutional weathering and erosion of material.

EROSION

The removal of material in dissolved or particulate form can occur through several processes. Dissolved removal occurs in water. Water carrying the dissolved material removes it from the slope through the ground or over the surface. The concentration of dissolved material (solutes) is generally highest in dry climates, but as there is little rainfall, the total amounts of the dissolved material removed from a location can often be less than in wetter areas. Once dissolved material is removed by water it generally travels far downstream.

In addition to dissolved material, water moves solid particles in what are called 'wash processes'. Rainsplash, rainwash and rillwash are the most important of these wash processes. The impact of raindrops can detach material, which then jumps into the air. The splash can cause the sediment to move either up or downslope but because of gravity there is an overall downslope movement in total. Ensuring there is a good vegetation cover to protect the soil surface from rainsplash forces is a good way of reducing erosion by rainsplash.

If raindrops land on flowing water moving over the land surface, then their direct impact on the soil surface is reduced. However, the flowing water itself can carry material. Where the flow over the surface is shallow, the combined effects of raindrop impact, which detaches sediment, and transport by flowing water over the surface are very effective and this combined process is called rainwash. When the water depth over the surface is deeper than 6 millimetres, raindrop detachment is weak and so the initial movement of a particle is more related to flowing water, in a process called rillwash. This erosion process is common in major storms.

Many poorly vegetated areas develop temporary rills, which are small erosion channels formed during storms. Wetting and drying, or freeze-thaw, accumulates material that infills the rills between storms. However, in a large storm, channels may form that are too large to be refilled before the next event and these are known as gullies. These channels collect water in subsequent events, rapidly enlarging the gullies further.

The wind can be an effective agent of erosion if the right type of sediment is available. Wind-blown sediment transport is called **aeolian transport**. Aeolian transport dominates in arid and semi-arid environments where there is little water. Strong winds are needed just to carry small particles. At typical wind speeds, medium-sized sand grains (up to 0.5 millimetres) are the largest grains that can be transported. However, the wind can carry finer dust thousands of kilometres and such processes are seen to be critical to global processes. The Sahara alone produces two thirds of the dust present in the Earth's atmosphere (260 million tonnes per year). Such dust has been shown to bring critical nutrients to the Amazon rainforest thousands of kilometres away. More locally, downstream deposition of wind-eroded material can lead to a range of landforms, including dune deposits. In coastal areas these may form individual features (see Chapter 5), but in dryland environments dunes tend to occur over wide areas known as sand seas. The Kalahari sand sea, for example, has an area of around 2.5 million square kilometres. About a fifth of dryland areas are covered by dunes. Many dunes may also occur in slightly more vegetated semi-arid and subhumid areas, and these may be relict landforms from past drier conditions at those locations. Sand dunes form when, for a given point, the input of sand is greater than the rate of loss. This accumulation might be related to the local topography when the moving sand meets an obstacle, such as a small hill, a plant or a rock. Once the sand accumulates, the small dune itself becomes the obstacle to flow, allowing further accumulation to take place. Normally, for dunes to prevail, there needs to be a dominant wind direction, or wind in a particular direction for a long duration, otherwise dunes may appear and then disappear as the wind direction changes. The shape of dunes formed depends on whether there is more than one prevailing wind direction, but the largest dunes can grow to around 100 metres in height. Once formed, dunes can, in turn, influence local wind velocities and therefore the erosion and deposition patterns in the area. Dunes tend to have a gentle gradient upwind side and a steeper downwind side. Large dunes may migrate over time across the landscape at rates of a few metres per year. Sand is transported by **saltation**, where the particles hop along the bed, and by rolling along the bed. Ripples across the sandy surface of dunes are commonly formed as the air flow rises and falls over small mounds of sand, forcing up sediment on the upwind side as the air rises and depositing the sediment on the downwind side of the ripple. These small ripples migrate over time as the sediment is moved and the features are short lived.

Wind erosional features in rock can also be formed because sand and dust particles in the air can be abrasive. Abrasion by sand is usually close to the ground (within 2 metres) as sand cannot be carried very high due to its large size. However, finer particles can also abrade at higher levels, smoothing rock surfaces and even whole hills, which are often known as **yardangs**.

Particulate material being eroded can do so as a large mass, rather than as individual particles. During mass movements (which can be very fast or very slow), a section of rock or soil moves together. Where large flows of water generate mass movement of sediment this tends to be faster than drier mass movements of sediment. The net effect of all forces operating on a mass of material controls when the material will move. Forces that encourage movement include gravity, water and wind. Flowing water can detach fragments of rock or soil if it passes over them rapidly by picking up (entraining) material from the surface, or by detachment of soil grains by raindrop impact. Friction and **cohesion** resist movement. Material begins to move when the forces promoting its movement become larger than the resistance forces. In simple terms, moving material will slow down and stop when it meets lower gradients, or where water carrying the material spreads out and moves more slowly or seeps into the ground. The **safety factor** is the ratio of the driving forces and resistance forces and is often carefully assessed when designing human-made slopes such as transport embankments or mining spoil heaps. The assessment tries to ensure that resistance forces will be greater than driving forces during different weather conditions.

There are many types of mass movements. In rapid mass movements there are slides, where the mass moves as a block, and flows, within which different parts of the material move over each other at different speeds. Fractures in rock provide weaknesses, which when fail enable slabs of rock to slide off downslope. Toppling can also occur when columns of rock become overhanging. Flows occur when there is more water mixed into the moving mass in proportion to the amount of sediment. In a slide there is little water within the moving material itself although water may have helped overcome friction to initiate the event.

There are also slow mass movements in which, as a whole unit, a mass of rock or soil moves slowly downslope. The slow mass movement of soil is called soil creep and typically operates at 1 to 5 millimetres per year. The movements are caused by expansion or contraction heaving (e.g. wetting and drying of soil or freeze-thaw action). Other movements are usually caused by biological activity, which mixes the soil in all directions. When this occurs on sloping ground gravity will result in more downhill movement than uphill and there is a gradual transport of material. Depending on the environment, one of the processes of creep may be more dominant. In cold, upland areas, freeze-thaw is probably the most important for creep processes, whereas in tropical forests biological mixing may dominate. Humans can accelerate soil creep through ploughing as each time the soil is turned over by the plough then soil moves. If the ploughing is directly up or downslope then there tends to be an overall downhill movement of soil when the soil settles back in the period after the ploughing. Where the ploughing direction follows the contour and cuts across the slope then material is moved either down or upslope depending on the direction of the plough. Where the plough turns soil downhill then this will result in an overall movement approximately a thousand times greater than soil creep. Contour ploughing in both directions as the plough moves one way across the slope and then the other still produces movement a hundred times greater than natural creep. Over the last few hundred years soil creep associated with ploughing may have been responsible for more soil movement in many areas than natural soil creep over the last 10 000 years.

Where there is lots of weathered material but the erosion process cannot carry it far (e.g. rain splash hitting the surface and moving particles) then the erosion process is transport limited. Situations in which there is not much sediment available to be transported but where the transport processes could, if more sediment was available, carry a lot more, are known as being in a supply limited condition. Places where transport limited conditions operate have a tendency to have a good cover of vegetation and soil, and over time the steepness of slopes often declines. Landscapes where removal of material is mainly supply limited tend to have little vegetation and soil and have steep slopes, which remain steep as they erode. The landscape can therefore contain many clues to the processes forming it. Convex hillslopes are associated with creep or rainsplash. Concave profiles are generally associated with rillwash. Mass movements generally lead to slopes with quite a uniform gradient except in situations of exceptional activity (e.g. coastal cliff erosion). Thus, landslides produce a landscape with uniform slopes. Semi-arid areas have stony, shallow soils with little vegetation; here hillslopes tend to be concave. Temperate, wet areas are generally dominated by creep, mass movements and solutional erosion under a dense vegetation cover and deep soils; here hillslopes tend to be convex. It is also possible to do your own assessment on smaller features to see what processes have been operating, for example by checking whether there are small mounds of sediment behind clumps of vegetation as an indicator of active wash processes.

Ice sheets, glaciers and rivers are active agents of erosion that shape landscapes over long time periods. These features and their associated processes are described in more detail in Chapter 5 and are therefore not covered further here.

SOILS

SOIL COMPOSITION AND FORMATION

Across the critical zone (see Chapter 1), the weathering processes described above result in rocks becoming fragmented, providing small habitats for plants which then add organic matter. This helps form soil. Soil is of crucial importance to humans. It acts as a zone for plant growth, which provides crops, supports animal life and holds water, influencing the amount and quality of water in rivers and lakes. Soil management may also be crucial in combating climate change because soils form an important store of carbon, which can be lost to the atmosphere if we are not careful with our landscape modifications and agricultural practices.

Soil is made up of minerals, organic matter, water and air. The amount of each of these components influences the properties of a soil. In most soils, the majority of the solid material is mineral matter which has been derived from rock weathering. Often only 2 to 6 per cent of the soil is organic matter, but it is still very important. Soil organic matter consists of decomposing plant and animal debris known as **litter**. It also consists of organic matter more resistant to decay known as **humus**, and living organisms and plant roots known as the soil **biomass**. Indeed, soil usually contains billions of bacteria in every handful. Litter is decomposed by soil organisms to produce humus. Plant nutrients, especially nitrogen, phosphorus and sulphur are released as litter breaks down, a process known as **mineralisation**. Soil organic matter holds mineral particles together, which stabilises the soil; improves the water-holding capacity of the soil; improves aeration; stores organic carbon; and is a major source of nutrients, important for soil fertility.

Air and water fill the gaps between solid soil particles. Soil air is very important as soil animals, plant roots and most microorganisms use oxygen and release carbon dioxide when they respire (breathe). In order to allow soil organisms to survive, oxygen needs to move into the soil and carbon dioxide must be able to move out of the soil. Therefore, the aeration of the soil is an important component influencing biological activity and litter decomposition.

Soil water contains dissolved substances, which are important for uptake by plant roots, and it helps to move dissolved chemicals through the soil (laterally, upwards and downwards) making them available for plants. Importantly, water is held in the soil despite gravity forces pulling it down. Even in very hot, dry deserts, some water is still found in the soil, showing that the forces holding water in the soil must be strong. This water remains in the soil because the combined chemical attraction of the water molecules to each other and the attraction of water to the soil particles is greater than the gravitational force. If you dip the end of the tissue into a bowl of water then you can watch the water being drawn up the tissue, showing that water does not simply just flow downhill. This is capillary action. Smaller pores exert stronger attraction forces on the water than larger pores. Therefore, capillary water will be driven to move from wetter parts of the soil to drier parts because the drier parts have more small pore spaces that are lacking water, thereby exerting a capillary attraction force on the water. This capillary action is also how plant roots draw water into the plant. If the soil is coarse, generally consisting of lots of large particles and large pore spaces between the particles, then it will not be able to hold as much water as a finer soil with smaller particles and smaller pore spaces. This pore water function explains why sandy soils cannot hold much water and are not as good at supporting plant growth as finer-textured soils that have more small pores.

Soil formation takes place over thousands of years, which is why mismanagement by humans that leads to large losses of soil by water and wind erosion is very problematic. The main input of soil material comes from the weathered rock below the soil. Mineral particles are released by weathering and contribute to the lower layers of the soil. Surface accumulation of organic matter from plants and animals is also important, as is dissolved material in water and particles carried by **precipitation** and the wind. The main losses of material from soils occur through wind and water erosion, plant uptake (but this usually gets returned back to the soil after death if the system does not have crop material removed from site) and **leaching**. Leaching is the removal of dissolved soil material. The leaching process is most rapid where there are large water inputs to the surface and where the soils are well drained (e.g. in an irrigated agricultural field with coarse soils and under-drainage installed). The percolating water carries dissolved substances downwards. While some of these substances are deposited in lower layers of the soil, some of the dissolved material may be completely washed out of the soil, potentially influencing groundwater or river water quality.

The factors that influence soil formation include climate, the ‘parent material’ (i.e. the weathered rock matter), the slope and organisms. The most influential factor is climate since it determines the moisture and temperature conditions for soil development; maps of major soil types often follow the climate zones. Soils in high latitudes are often very shallow and develop slowly, while soils that are several metres deep are typical of tropical areas. Parent material influences soil formation through the influence of weathered material on soil processes, while slope steepness, aspect and altitude all affect the local climate as well as drainage and erosion conditions. The type of vegetation influences the type and amount of litter that is returned to the soil, while different soil types support different vegetation communities. For example, in conifer forests there will be a deep litter layer of thin waxy needles which only decompose slowly. Vegetation also protects the soil from water and wind erosion by intercepting rainfall, decreasing the role of rainsplash (see section on erosion above).

Soils are often classified by the nature of their vertical profile. If you dig a hole through a typical soil, from the surface to the bedrock, the soil will be made up of a series of horizontal layers known as **soil horizons** (Figure 2.6). In some soils the horizons

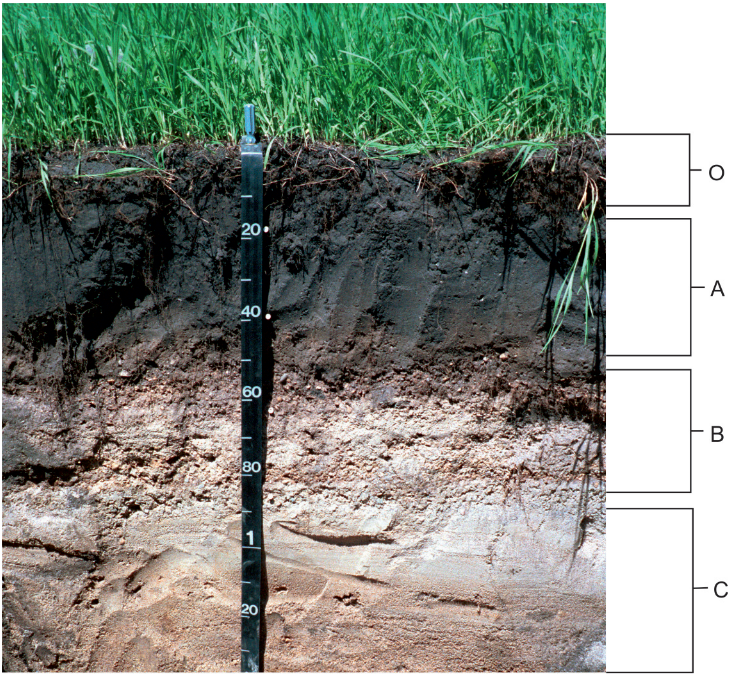

Figure 2.6 A soil profile from northern Minnesota.
(Source: courtesy of the USDA Agricultural Research Service, horizon labels have been added to the original photo.)

have very distinctive colours and clear boundaries, while in others the changes with depth may be gradual with rather uncertain boundaries. Soil horizons are given letters according to their formation and their relative position in the profile. The uppermost O horizon is dominated by an accumulation of fresh and partially decomposing organic matter. Below this, the dark A horizon contains humus and minerals. The typically pale E horizon is dominated by leaching or removal of material by plant roots (not present in the soil shown in Figure 2.6), while the underlying B horizon is often a zone of accumulation of matter from leaching above and weathering below. There is usually a transition from the B horizon into the C horizon, which is mainly weathered parent material known as the **regolith**. The bedrock at the base of the soil profile is typically designated as the R horizon. Not all soils contain all of the layers described above.

Laterisation (also known as ferralitisation) occurs in tropical soils where high temperatures and plentiful rainfall result in fast rates of both weathering and leaching so that there are horizons depleted in bases (e.g. calcium, magnesium, potassium and sodium) and enriched in silica and oxides of aluminium and iron. 'Podzolisation' (producing soils called podzols) may occur where there is strong leaching, often where there is plentiful rainfall and good drainage, typically under forests or heaths. Organic acids from litter are washed through the soil and react with iron and aluminium compounds that are transported downwards from the E horizon by percolating water and deposited in the B horizon. Podzols tend to be rather unproductive for agriculture as fertilisers are readily washed away and the soils are acidic. In waterlogged conditions there are reactions with iron products in the soil driven by microorganisms in a process called **gleying**. The soil becomes grey or bluish. In arid and semi-arid areas, water is drawn upwards to the soil surface and as the water evaporates salts are left behind at or near the surface. Thus, a wide range of soil types and conditions exist.

There are several different soil classification systems and each system contains around 15 to 30 different soil types, often with subcategories for main types. The terminology used in one classification can be different for the same soil in another classification scheme, which can cause some confusion. While there are too many examples to describe in this book, some soil types from the US Department of Agriculture system are: vertisols (swelling clay soils with deep, wide cracks), histosols (organic rich soils), andisols (formed on volcanic parent material, especially ash) and oxisols (red/yellow/grey soils of tropical and subtropical regions with strongly weathered horizons enriched in silica, clay and oxides of aluminium and iron, which are acidic with low nutrient status).

PHYSICAL PROPERTIES OF SOILS

Soil texture and structure influence how soils work and how they can be managed. They control the water holding capacity and therefore a soil's ability to support plant growth. They also control its permeability and hence the soil's role in controlling rates of water flow through the landscape to rivers and aquifers (see Chapter 5).

Soil texture refers to the relative proportions of sand-, silt- and clay-sized materials. Clay particles are smaller than 2 micrometres (two-millionths of a metre), silt is between 2 and 60 micrometres and sand is between 60 and 2 000 micrometres in diameter. The texture controls the water holding capacity, aeration, drainage rate, organic matter decomposition rates, compaction, susceptibility to water erosion, ability to hold nutrients and leaching of pollutants. Figure 2.7 shows how soil texture classifications are based on the relative proportion of each particle size. For example, reading Figure 2.7 shows that if a soil is 40 per cent sand, 30 per cent silt and 30 per cent clay it would be classified as a clay loam.

Good soil structure is important for achieving well-drained and aerated soils. Soil structure refers to the arrangement of soil particles. Soil particles usually attach themselves to each other in formations called **peds**. Where a soil lacks peds, such as in a sand dune, the soil is described as structureless. Soil structure is characterised in terms of the shape, size and distinctiveness of these peds with four principal types: blocky (roughly equally sized on each side, almost cube-shaped, but the peds can be angular or more rounded), spheroidal (sphere-shaped), platy (horizontal plates) and prismatic (vertically elongated columns of soil with flat face). Clay particles and organic compounds largely hold peds together. As a

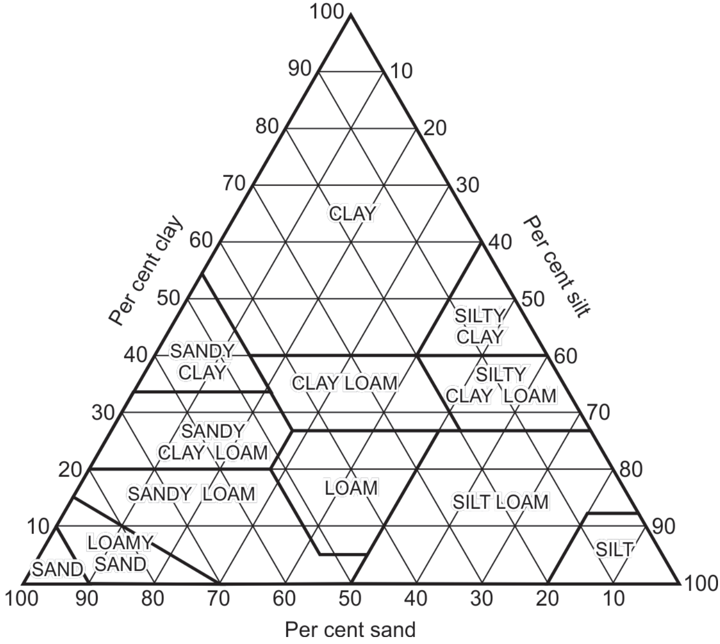

Figure 2.7 A common classification for soil texture.

result, coarse-textured soils tend to have weakly developed structures, whereas fine-textured soils generally have moderate to strong structures. The strength of peds influences resistance to erosion and the ease of cultivation.

CHEMICAL PROPERTIES OF SOILS

Soil chemical properties are strongly influenced by parent material and organic matter content, which provides clay and organic particles. Clay minerals are formed from the weathering products of aluminium and silicate minerals. As clay minerals are small, a volume of clay will have a large total surface area around all of its particles compared to the same volume of sand. An ion is an atom or group of atoms with an electrical charge (either positive or negative). Clay particles have a negative electrical charge so these ions can attract and hold water and cations (positively charged particles). Therefore, they have a fundamental influence on both the physical and chemical properties of the soil. The concept of the **cation exchange capacity** is an important one. It is essentially a measure of the soil's ability to hold and release various elements such as plant nutrients. Experiments during the nineteenth century showed that when ammonium chloride was added to the top of a soil (as part of nitrogen fertiliser), a solution of calcium chloride came out of the bottom of the soil. The cations of ammonium and calcium were rapidly exchanged in this process. The process is also reversible. The negative charge on the clay and organic humus particles is balanced by positively charged cations, which are attracted to the clay and humus particles. These cations are referred to as exchangeable cations because cations in the soil water solution can displace the cations on the clay surface. This exchange between a cation in soil water solution and another on the surface of a clay particle is cation exchange. Cation exchanges are balanced reactions, so if an ion with two positive charges, such as calcium (Ca^{2+}), is washed through by a solution of sodium (which has a single positive charge; Na^{+}), then it will take two sodium ions to replace one calcium ion. The cation exchange capacity is the ability of an amount of soil to hold cations, and this depends on the overall negative charge of the clay particles present. The cation exchange property controls fertility and acidity and also means that soils act as an important buffer between the atmosphere and groundwater, thereby potentially reducing pollution of water courses.

The acidity of a soil is important since it affects many soil processes, the plants that grow and what happens to some pollutants. Many polluting heavy metals become more soluble in water under acid conditions and can then move downwards with water through the soil to groundwater or river water. The concentration of hydrogen ions in soil solution determines whether a soil is acidic, neutral or alkaline. Concentrations of these ions are very small and so the pH system was developed. Low numbers (starting at 1) on the pH scale are acidic, 7 is neutral and large numbers, up to 14, are alkaline. In the pH scale a change of one unit represents a 10-fold change in hydrogen concentration. So, a pH of 5 means the soil solution has 10 times the concentration of hydrogen ions than at a pH of 6. Most soils have a pH of between 3.5 and 9, and very low values are often associated with soils rich in organic matter, such as peat.

Green plants cannot grow properly without 16 essential elements available in the correct proportions. The availability of these essential nutrients for plants is influenced by soil pH. On the basis of their concentration in plants, the 16 elements are divided into macronutrients (carbon, oxygen, hydrogen, nitrogen, phosphorus, sulphur, calcium, magnesium, potassium and chloride) and micronutrients (iron, manganese, zinc, copper, boron and molybdenum). As a comparison, the typical amount of potassium in a soil is 1.5 per cent of the total mass whereas molybdenum only accounts for a hundred thousandth of a per cent. A pH range of 6 to 7 is generally best for plant growth as most plant nutrients are readily available at suitable concentrations in this range. High soil pH results in phosphorus and boron becoming insoluble and unavailable to plants. Most nutrients are more soluble in low pH (acidic) soils, which can result in high or toxic concentrations of them. On the other hand, phosphorus and molybdenum become insoluble at low pH, and so become unavailable to plants.

HUMANS AND SOIL

Human activity can alter the soil. It is astonishing to note that the total global area of soil degraded by humans (over 20 million square kilometres) exceeds the amount currently being used for farmland. This degradation is due to deforestation, overgrazing and poor agricultural management. Degradation includes soil erosion, soil acidification, soil pollution, the reduction of organic matter content and **salinisation.**

Over the last half century, it is thought that between one third and a half of the world's arable land has been lost by erosion. While erosion is a natural process, humans have dramatically accelerated this process. Farmland is often left bare or has a low vegetation cover for a considerable period of the year, and old obstacles to erosion like large woody plants have often been removed. This makes it easier for the wind or wash processes to transport surface sediment. Tractor wheels also compact the soil, making fast channels for water flow over the surface that often accelerate erosion. Accelerated soil erosion results in a reduction of soil depth because soil is lost at a faster rate than it is being formed, often with the more organic, rich upper layers being lost first. The erosion can sometimes lead to large gullies forming, making it difficult to farm. The sediment lost through erosion moves downstream and can often silt-up roads, reservoirs and water courses. Some agricultural chemicals are also bound to the sediment, polluting water courses. Erosion control often involves reducing grazing intensity, planting strips of vegetation to form barriers to reduce the wind and catch sediment, the use of cover crops when the surface would normally be left bare, more careful ploughing and building terraces to reduce the slope and catch any sediment (so the landscape looks like a series of steps coming downhill).

Salinization occurs when soluble salts of sodium, magnesium and calcium accumulate so that soil fertility is reduced. This is mainly a problem in warm, dry regions where evaporation and upward movement of water in soils exceeds downward movement from rainfall and percolation. Furthermore, irrigation of the land with water of a high salt content (i.e. where a lot has evaporated before being used so that the concentration of salts in the water is greater) worsens the situation. In some countries around 10 per cent of their arable land is affected by salinization.

Soil pH has been reduced by the burning of fossil fuels, which has resulted in rainwater becoming more acidic. Also, the harvesting of crops and overuse use of nitrogen fertilisers causes acidification. Soil acidification increases the solubility of heavy metals (metallic elements with a density greater than 6 grams per cubic centimetre), which can be toxic to plants, reducing growth rates or altering the types of plants for which the soil is suitable (e.g. forest decline in central Europe). The soil organisms may also be affected, with species changing towards those which are more tolerant of acidic conditions, resulting in a slower rate of litter decomposition.

Heavy metals such as copper, lead, zinc and mercury are present naturally in soil, but atmospheric pollution and application of sewage sludge, farm wastes and leaching from landfill sites can add heavy metals to soil. The worst areas for heavy metal pollution are those around industrialised regions, such as in northwest Europe. Mining, smelting, energy generation, agriculture and the wear of vehicles and other machinery are sources of heavy metal pollution to soils. Heavy metals build up in soil as they become bound to organic matter and clay minerals and are generally not taken up by plants. However, if the soil becomes more acidic this can result in heavy metals being released into soil water and becoming available for plants to take up, or for leaching into rivers, lakes and groundwaters. This is the crucial phase because, if this happens, then the crops we eat may have toxic levels of heavy metals or water may become dangerous to drink.

Modern agriculture relies on pesticides for crop protection and disease control, and fertilisers to provide additional nutrients, such as nitrogen, phosphorus and potassium, to maintain good plant productivity. Pesticides can often attack living things that they were not intended for and can leach into groundwaters and rivers. Fertiliser use has increased tenfold over the last 60 years. This is fine as long as it has been applied in the correct amount for the soil and crop and at the right time of year so that it is not wasted and simply washed out of the soil. However, river water, groundwater and lakes have seen increased nitrate concentrations related to the leaching of fertilisers. Nitrates in drinking water are a health hazard and can damage river- and lake-based plants and animals. **Organic farming** relies on biological processes for crop and livestock production rather than the use of manufactured pesticides and fertilisers. Practices such as crop rotations, planting varieties more resistant to disease, using plants that take nitrogen out of the atmosphere and applying compost and manures are common in organic farming.

The above human impacts on soils have led to concerted efforts, legislation and policy to protect and restore soils as an important underpinning resource for human life. However, much more effort is needed, particularly as the world's population continues to grow, placing more demand on our soil resource. At the Paris climate conference of the United Nations Framework Convention on Climate Change in 2015, an initiative was launched to target increased soil carbon by 0.4 per cent per year ('4 per mille initiative'). If achieved, this could be important for drawing carbon dioxide from the atmosphere to mitigate climate change and reversing the losses in soil carbon caused by intensive agriculture over the past few decades. However, there is some scepticism about whether such ambitions can be achieved and, of course, not all soils are alike. As part of climate change policy development (achieving 'net zero' carbon emissions by 2050 – see Chapter 4), countries such as the UK are seeking ways in which the agriculture and land-use sector can contribute, including encouraging increased carbon uptake into soils. Planting more hedgerows and trees on farmland, keeping stubble on the field rather than burning it, using less disturbing ploughing techniques (or none), increasing the proportion of grass cover, and including ‘rotations’ where fields are left without cropping for some years to recover, can all allow soil organic matter to accumulate. The latter strategy can be particularly useful if the fallow years involve the fields being planted with grass and clover-type vegetation that capture good nutrients from the atmosphere. Furthermore, these sorts of techniques can make the soil system more resilient to extreme weather events such as droughts, thereby helping to mitigate climate change impacts. Soils with good structure and plentiful organic matter can hold more moisture during dry weather. In addition, some forms of precision agriculture, supported by large datasets, can contribute to increased carbon storage in soils and improved soil functions while also allowing good crop yields. Box 2.2 provides further detail.

BOX 2.2 BIG DATA TO SAVE SOILS

New techniques that allow us to compile, process and analyse vast datasets (‘big data’), including artificial intelligence, open up opportunities for ‘smart’ soil management. Over recent years, more farmers have been adopting precision agriculture based on location data from their farm vehicles, soil sensor data, crop conditions, weather data and so on. This allows them to plant, fertilise, water, plough and administer pesticides at carefully controlled rates depending on their location within a field, soil conditions and so on. By feeding back data from each sensor and each vehicle on each farm in real time into large data storage and processing facilities, it is possible to use computers to ‘learn’ how key factors affect soil conditions or crop yields and, therefore, continuously improve precision agriculture to support soil conditions and wider benefits. In the machine-learning approach, computer systems are trained to identify the (sometimes hidden) patterns based on available data, such that accurate future forecasts can be made about the best locations or timings to apply different farm management practices. The more data fed into the computer system, the more the machine can learn, process and improve forecasts and the better our understanding of spatial patterns of soil types and conditions will be.

For example, smart camera sensors on farm machinery can detect plant and soil characteristics. Machine learning algorithms can then process these data to determine very spatially targeted optimum chemical doses for crops. This has resulted in reported 80 to 90 per cent reductions in uses of chemicals at farms that have trialled such technology, and these reductions are now being promoted by companies such as John Deere who sell farm equipment and data support services. These advances save farmers money in chemical use and also provide environmental improvements.

SUMMARY

- The Earth's crust is formed of moving plates; at the edges of plates earthquakes and volcanic activity are common.
- As continents collide mountains form.
- In the centre of oceans, new crust is formed; at the edge of oceans, the crust sinks back into the mantle. Ocean crust is relatively young, being less than 200 million years old, whereas continental rocks can be billions of years old.
- Weathering by physical and chemical processes wears down rock. Climate and rock type are important controls of weathering process rates.
- Erosion transports weathered material by water, wind, and slow and fast mass movement.
- Soil is made up of minerals from weathered rock, organic matter, water and air.
- Soil formation is affected by climate, the parent material, topography and organisms.
- The texture of soil particles, structure and chemistry of a soil are crucial in determining its water and nutrient exchange capacity and hence its use for plant growth.
- Careful soil management is required as humans have degraded large areas of soil across the Earth through poor agricultural practice and pollution.
- Good soil management can reduce climate change impacts by storing more carbon on land, saving energy used to produce chemical fertilisers and pesticides, and making the soil system more resilient to extreme weather events.

FURTHER READING

Ballantyne, C.K. (2018) *Periglacial geomorphology*. Chichester: Wiley-Blackwell.

See, in particular, Chapters 10 and 11 on rock weathering and mass movements.

Brady, N.C. and Weil, R.R. (2016) *The nature and properties of soils* (15th edition). Harlow: Pearson Education.

This is a popular soils textbook; many of its examples are North American.

Gregory, K. J. (2010) *The Earth's land surface*. London: Sage.

A very easy to use textbook with useful case studies and comprehensive coverage of processes, including chapters on each main global environmental type.

Grotzinger, J. and Jordan, T.H. (2020) *Understanding Earth* (8th edition). Boston: Bedford/St Martin's.

This is a full colour text with additional media materials to explain the internal workings of the Earth, especially those around plate tectonics, earthquakes and volcanism.

Holden, J. (ed) (2017) *An introduction to physical geography and the environment* (4th edition). Harlow: Pearson Education.

A textbook with expert contributors providing more in-depth material on all of the topics covered above, including chapters on 'Earth geology and tectonics' (pp. 29–52), 'Weathering' (pp. 347–376), 'Sediments and sedimentation' (pp. 407–428), 'Slope processes and landform evolution' (pp. 377–406) and 'Soils' (pp. 429–464).

Huggett, R.J. (2016) *Fundamentals of geomorphology* (4th edition). London: Routledge.

This university textbook covers tectonic processes, weathering, erosion, and other processes and landforms also relevant to Chapter 5.

3

ATMOSPHERE, OCEANS, CLIMATE AND WEATHER

THE ATMOSPHERIC BLANKET

The Earth is surrounded by a layer of gas which extends around 500 kilometres above the Earth's surface. However, near the top of the atmosphere the density of gas is very small, with three quarters of the gas instead concentrated within the lowest 11 kilometres next to the Earth's surface. The lower layer of the atmosphere is called the **troposphere** and extends to about 6 kilometres altitude over the poles and about 15 kilometres over tropical regions. The troposphere is where air mixes most rapidly and where we experience weather.

The Earth's atmosphere acts as a filter protecting us from space debris and harmful radiation. The Earth receives only two-billionths of the Sun's total energy release, but this energy is the main driver for water, air and wind motions, and most life on Earth. As Table 3.1 shows, only around half of the Sun's energy that reaches the Earth's atmosphere makes it all the way down to warm the Earth's surface. Some of this energy reaching the Earth's surface is used for processes, such as the evaporation of water or for plant growth. However, most absorbed radiation from the Sun (known as short-wave radiation) is transformed by the land, oceans and vegetation and emitted back into the atmosphere as long-wave radiation in the form of heat energy (invisible **infrared radiation**). Except for the 18 per cent of incoming solar energy that is temporarily absorbed (Table 3.1), the atmosphere is mostly transparent to incoming short-wave radiation. This means, perhaps surprisingly to many people, that the air in the troposphere is mainly heated

DOI: 10.4324/9781003124405-3

Table 3.1 Fate of the Sun's energy reaching the Earth

Fate	*Percentage of Sun's energy reaching Earth (%)*
Scattered and reflected by clouds	21
Scattered and returned back to space by the atmosphere	6
Absorbed by atmosphere and clouds temporarily before being returned back to space	18
Reflected back to space by surface features such as ice, snow etc	4
Absorbed by the Earth's surface	51

from below by long-wave heat energy emitted by the Earth's surface. Thus, the atmosphere should be warmer close to the Earth's surface but cool with altitude in the troposphere. Less dense gases will naturally seek to rise and more dense fluids will seek to fall. Since the air is warmed by the surface below during the day, **convection** occurs whereby the less dense air near the surface seeks to rise above cooler, denser air, which in turn sinks towards the Earth's surface. As the air rises, it cools because it is able to expand due to the lower air pressure at higher altitudes (as the pressure of a gas decreases the temperature will decrease). The result of these processes is that there is large-scale vertical mixing of the air within the troposphere as rising warm air is replaced by cooler descending air.

The atmosphere is made up of mainly nitrogen (78 per cent) and oxygen (21 per cent). The remaining one per cent is made up mainly of argon. There are also small concentrations of other gases, such as hydrogen, water vapour (the gaseous form of water), methane, nitrous oxide, ozone and carbon dioxide. Despite their low concentrations, some of these other gases are important for the climate we experience. While the gases of the atmosphere are almost unaffected by the short-wave radiation provided by the Sun, some of them readily absorb long-wave radiation produced by the Earth's surface. Unlike oxygen and nitrogen, some gases, such as carbon dioxide, methane, water vapour and nitrous oxide, absorb the thermal energy emitted by the Earth's surface and provide a form of 'blanket' over the Earth. They radiate this energy back down to Earth again, which in turn is absorbed by the Earth, further enhancing the heating of the atmosphere. A greenhouse does a similar thing. The glass allows short-wave radiation to pass through it to the soil and plants, which then absorb the radiation and re-radiate thermal energy back towards the glass. However, the glass traps the long wavelength heat energy and the warmer air inside the greenhouse. The **natural greenhouse effect** in the atmosphere is a good thing. If this did not happen, then during the day the Sun's energy would be absorbed by the land, oceans and vegetation at the surface and then transformed into heat, which would be radiated back into the atmosphere. However, at night all of this energy would radiate back into space and so the Earth's surface temperature would fall to extremely cold levels very quickly. The greenhouse gases prevent this from happening by retaining some of the energy within the troposphere, delaying its release back out to space and keeping the planet at a good temperature for life. The average temperature of the Earth's surface is 15°C, but without the natural greenhouse effect the average temperature across the Earth would be around −20°C.

The composition of the atmosphere has changed through time. The Earth is around 4.6 billion years old. The early atmosphere mainly consisted of nitrogen gas and carbon dioxide, with no oxygen gas. Oxygen gas did not start to appear in the atmosphere until about 2 billion years ago. It was at this point that bacteria evolved, and they functioned by absorbing carbon dioxide from the atmosphere and then releasing oxygen through **photosynthesis**. More recently, humans have also changed the composition of the atmosphere through the burning of fossil fuels and release of other chemicals, a topic further explored in Chapter 4.

LARGE-SCALE ATMOSPHERIC CIRCULATION

The rate of temperature change with altitude is known as the **lapse rate**. The normal rate at which temperature declines with altitude in the troposphere is around 6.4°C per kilometre (this is variable) and is known as the **environmental lapse rate**. However, a rapid temperature change associated with a rising and expanding parcel of air is described as ‘adiabatic’, meaning that there is no interchange of heat between the rising air parcel and its surroundings; the temperature change is internal to the air parcel. In this instance the rate of temperature decrease with altitude is known as the **dry adiabatic lapse rate** and is 9.8°C per kilometre. Cooling of a rising air parcel may result in the air becoming saturated with water vapour (the **dew point**), because air can hold less water vapour at colder temperatures. This will lead to condensation of water droplets and the formation of clouds and precipitation. (Think of a cold drink you take out of the fridge and place in a warm room; water droplets form on the outside of the drink container as the air around it has cooled to the dew point and becomes fully saturated, leading to condensation of water vapour onto the container.) When water vapour (gas) condenses into liquid water it releases heat, which then warms the air slightly. Therefore, the lapse rate within this air parcel is less than the dry adiabatic lapse rate and is known as the **saturated adiabatic lapse rate**. The exact value of this lapse rate varies depending on the amount of moisture in the air and the temperature. The additional heat energy released by the condensation process may force the air parcel even higher to form large clouds of condensed water. So, it is the difference between the lapse rates that determines whether there will be a continual rise of the air mass and cloud formation or whether there will be very stable conditions (i.e. when the environmental lapse rate is less than the dry and saturated adiabatic lapse rates).

The above processes highlight the important role of water vapour in atmospheric motion. On average, the atmosphere holds about a fortieth of global annual rainfall (about 25 millimetres depth of water if it were all deposited evenly across the whole Earth). Regular evaporation from land and oceans maintains rainfall throughout the year in different parts of the world. This is not evenly spread across the planet and some areas provide a lot more evaporation than they receive as rainfall, and vice versa, indicating that water moves significant distances in the atmosphere. As we will see later on, evaporation is also important for controlling major ocean circulations, which have a big impact on the Earth’s climate.

The above description of vertical atmospheric processes does not explain why winds and water vapour move horizontally across the Earth or why the same location can experience periods of calm and periods of storminess. The Earth's wind circulation patterns help form the climatic zones. There are two main processes that drive global wind circulation. The first is the uneven distribution of the Sun's radiation over the Earth's surface due to its spherical shape. Figure 3.1 shows how the same amount of solar radiation is spread over a large area near the poles compared to the equatorial regions where it is more concentrated. This creates a north–south temperature gradient, and as heat is always transferred from hot materials to cooler materials, the warm air (and oceanic water) from the equator will naturally try to rise and move polewards at high altitudes within the troposphere to be replaced by lower-level winds/ocean currents moving in the opposite direction.

The second driver of global wind circulation is the Earth's rotation. If the Earth did not rotate and only one side faced the Sun, surface winds would blow from the cold dark side to the hot daylight side, as rising air over the hot side would need to be replaced by colder air drawn in. However, the Earth's rotation acts to create an apparent deflection of winds to the right in the northern hemisphere and to the left in the southern hemisphere, which is a process known as the **Coriolis effect** and which is stronger as you move toward the poles.

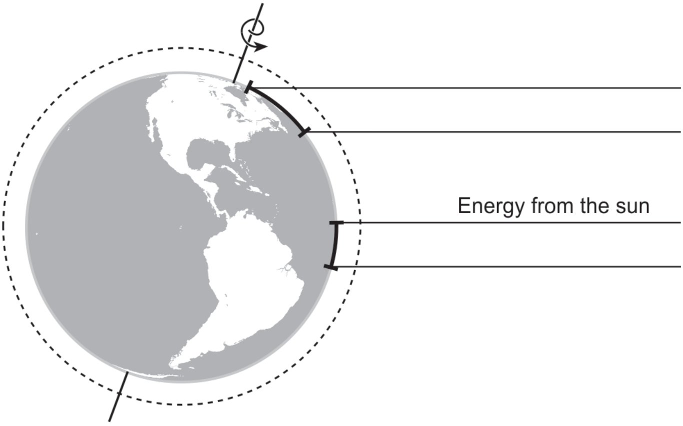

Figure 3.1 The Sun's energy is more concentrated near the equator and more diffuse at the poles.

The other effect of the Earth's rotation relates to the tilt of this rotation and its impact on the seasons. The rotation of the Earth is not perpendicular to its rotation around the Sun but is tilted to 23.5°, as shown in Figure 3.1. This drives seasonal cycles upon Earth because the Sun appears overhead at midday at the Tropic of Cancer (23°27′ N) on 20–21 June and overhead at midday at the Tropic of Capricorn (23°27′ S) on 21–22 December. Areas that are polewards of the Arctic Circle (66°33′ N) have at least one full 24-hour period of daylight on 20–21 June, with the same being true for the Antarctic Circle on 21–22 December. There is almost six months of darkness in the winter at the poles and six months of summer daylight. Thus, the northern hemisphere receives more of the Sun's energy from the March equinox (21/22 March) to the September equinox (21/22 September) than the southern hemisphere, with the reverse true for the other half of the year. The midday Sun is directly overhead at the equator during equinoxes and on these days the length of day and night across the Earth is the same in all places.

The Coriolis effect combined with the latitudinal temperature gradients results in large atmospheric circulation cells, as shown in Figure 3.2, with zones of rising and falling air creating low and high pressure at the surface. The Hadley cells are formed by hot rising air near the equator that flows towards the poles and then sinks at about 30° north and south before returning at low levels back to the equatorial regions. As air sinks (and thereby warms), creating high pressure at around 20 to 30° latitude, moisture condensation is not common, and so this region consists of clear skies and light winds. It is at this high-pressure zone that most of the world's deserts are found. Between 30° north and south there are equatorward-flowing easterly winds (i.e. winds moving from east to west) known as the **trade winds**. These meet from north and south near the equator at a zone of low pressure caused by rising warm air. This zone is known by atmospheric scientists as the **intertropical convergence zone**. Here conditions are favourable for warm, moist, rising air, condensation of water vapour, cloudiness and large amounts of rainfall. The intertropical convergence zone is only a few hundred kilometres wide.

Over the poles there are similar circulation cells to the Hadley cells but with descending cold air at the poles and air flowing

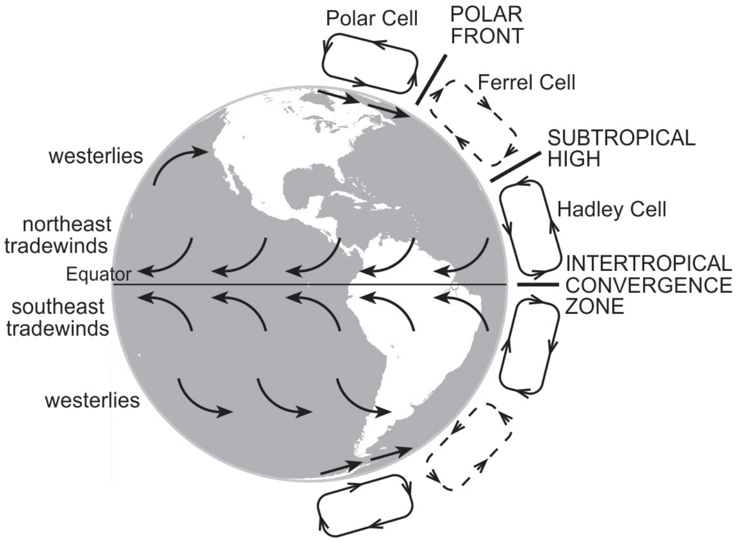

Figure 3.2 Schematic representation of the major atmospheric circulation cells and surface wind directions.

towards the equator at the surface. However, the Coriolis effect diverts the surface polar winds in an easterly direction, as shown by the arrows on the schematic diagram of the Earth in Figure 3.2. The polar cells tend to be weak as the Sun's energy is less intense at high latitudes.

Between the Hadley cell and the polar cell is the Ferrel cell. The Ferrel cell consists of sinking air at around 30° north and south and poleward-moving westerly winds at the Earth's surface. Rising air occurs at the boundary between the Ferrel cell and the polar cell, and this is called the **Polar front**. At the surface within the Ferrel cell are westerly winds carrying eastward-moving cyclonic (anticlockwise) and anticyclonic (clockwise) circulation systems.

At higher levels in the troposphere the belt of prevailing westerly winds associated with the Ferrel cell is disturbed by large undulations. This leads to a wave formation that loops around the Earth known as **Rossby waves** (Figure 3.3), and is a bit like icing poured over the top of a spherical chocolate pudding. The location of Rossby waves is affected by large mountain ranges so that the

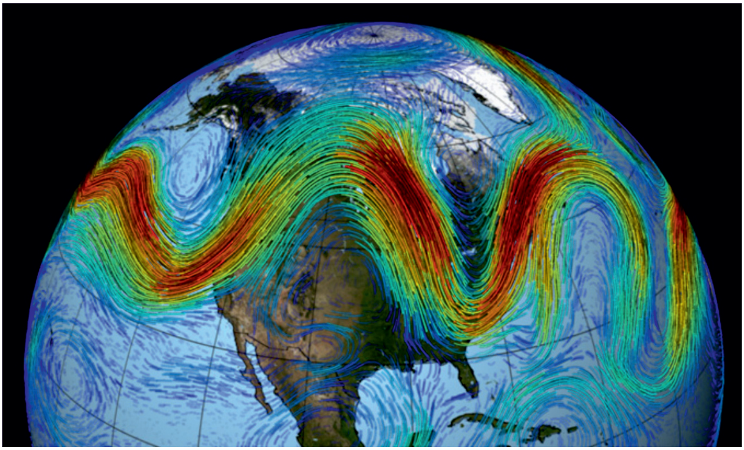

Figure 3.3 Rossby waves at one point in time in the atmosphere above North America, the North Pacific and North Atlantic, with faster moving regions of air shown in red and slower moving regions in light blue.

(Source: NASA's Goddard Space Flight Center, https://oceanservice.noaa.gov/facts/rossby-wave.html.)

waves, particularly in the northern hemisphere (where there is more land), are locked into preferred locations. The air is colder towards the poleward side of the Rossby waves and much warmer on the equatorward side. In the winter there tends to be two troughs in the waves near the eastern edges of North America and Asia, with ridges over the Pacific and Atlantic. This corresponds to troughs being associated with cold air over the winter continental land masses and ridges over the warmer oceans.

Within the Rossby waves are fast moving bands of air (at least 30 metres per second) called **jet streams**, which are caused by sharp temperature gradients. Jet streams tend to be thousands of kilometres long, hundreds of kilometres wide and several kilometres deep. Aircraft can make use of the jet streams by travelling within them when moving west but avoiding them when moving east. For instance, the flight from the west coast of the USA to Europe can be an hour shorter than the reverse trip. The shape of the Rossby waves and the location of jet streams are used in weather forecasting as they are associated with the formation of large circulating air masses across the Earth's surface in the mid-latitudes. These circulating surface air patterns bring warm and cold air masses together and can lead to different calm/windy or dry/wet conditions. There are several online animations of the Rossby waves (e.g. https://oceanservice.noaa.gov/facts/rossby-wave.html) showing the wind speeds associated with them and how the waves move over time, with eddies that may break away to form rotating storms.

LARGE-SCALE OCEANIC CIRCULATION

The oceans cover 71 per cent of the Earth. The interactions between the oceans and the atmosphere play a major role in controlling the climate and weather at the Earth's surface. There are four primary factors that control ocean circulation: i) water **salinity**; ii) water temperature; iii) surface winds; and iv) the Coriolis effect.

i) Contributing to water salinity, chemical weathering of rocks on land (see Chapter 2) produces dissolved materials that enter the oceans. Some of these dissolved chemicals undergo reactions within the ocean to take them out of the water and deposit them on the ocean floor. For example, sea creatures extract dissolved calcium from the water to build their shells and then when they die, the calcium drifts down onto the ocean floor. The balance of inputs from rivers and losses onto the ocean floor largely controls the chemical composition of the oceans. The concentration of salts in the ocean therefore varies depending on location and time. Surface water salinity can be diluted by melting ice or rainwater, or it can be concentrated by evaporation. The salinity of the ocean water around the dry subtropics of 20 to 30° north or south tends to be greater than elsewhere because evaporation is greatest here.
ii) In terms of water temperature, the surface of the ocean absorbs the Sun's energy and gains heat. More heat is gained than lost in the low latitudes and more is lost than gained in the high latitudes. Just as with the atmosphere there is a tendency for warm surface water to move polewards. This transport of heat from the equator towards higher latitudes

provides good regulation of the Earth's climate system, otherwise the poles would be even colder and the tropics even hotter than they are at present. Water provides an excellent long-term store of heat for the Earth as it takes more energy to heat water by 1°C (and more energy is released when water cools by 1°C) than for any other common substance.

Both salinity and temperature control the density of seawater. If the density of seawater increases with depth, then the water is said to be vertically stable. If, however, there is more dense water on top of less dense water then vertical mixing of water will take place. This means that if there is a warm, strong wind encouraging evaporation at the ocean surface then this will lead to more salty and dense surface waters, leading to instability and mixing of water.

iii) The surface currents of the ocean, as shown in Figure 3.4, are driven by the surface winds. For example, the trade winds (shown in Figure 3.2) drive the northern and southern equatorial currents, moving in a westerly direction parallel to the equator.

iv) The surface currents are deflected by the continents and the Coriolis effect (to the right in the northern hemisphere and to the left in the southern hemisphere). This creates warm currents along the eastern coasts of the Americas, Australia, Asia and Africa. In the North Atlantic this warm current is called the Gulf Stream, which brings warm conditions to northwest Europe. The Gulf Stream forms the western and northern parts of the North Atlantic subtropical **gyre**. The five subtropical gyres (see Figure 3.4) are the most dominant surface features of the world's oceans, with their centres located at 30° north or south. The centres of the gyres were often avoided by sailors of the past due to their calm wind and calm ocean current conditions, which would mean they could not make progress very quickly. Interestingly, the centres of these gyres have become collecting areas for the floating rubbish we have discarded into the world's oceans (Box 3.1). In each ocean there is a westerly current in the temperate mid-latitudes with a cold current flowing back towards the equator on the western edge of the continents. There is also a

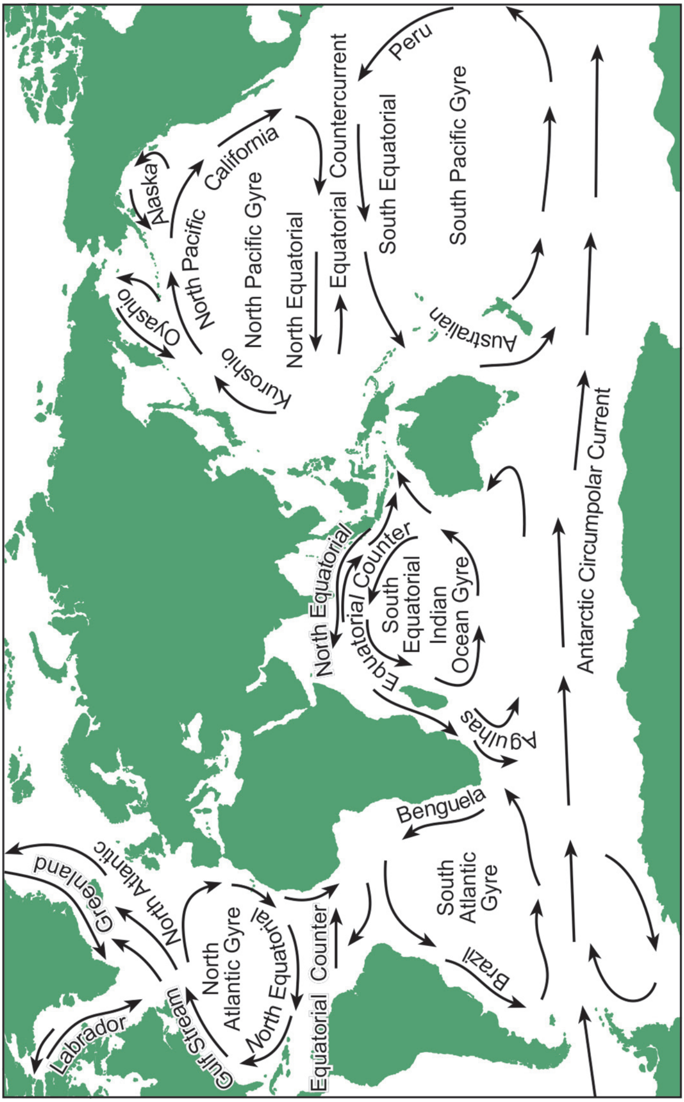

Figure 3.4 A map of the main surface ocean currents.

circumpolar current around Antarctica around 60° south. This is not mirrored around the Arctic as there are too many land masses in the region acting as a barrier to water flow.

BOX 3.1 THE OCEANIC GARBAGE PATCHES

Plastic waste can be harmful to marine life, causing entanglement, choking, starvation when plastic fills stomachs and the release of toxins. Around 400 million tonnes of plastic are produced annually, but unfortunately much of this ends up as discarded waste (packaging or unwanted, used or broken items). Several million tonnes of this waste enter the oceans each year, washed out by rivers or discarded in coastal areas.

The ocean currents transport floating plastic debris, which can be found in all of the world's oceans, including remote areas of the Arctic and Antarctic Oceans. However, the largest concentrations of oceanic plastic have been reported from the calm centres of the subtropical oceanic gyres, which are often very remote from large land masses. Here, researchers have found hundreds of kilograms of plastic per square kilometre, often many decades old, and over a million pieces of plastic greater than 0.5 millimetre in diameter per square kilometre. Hence these areas have been described as ocean garbage patches.

Plastic can slowly break down into smaller pieces under sunlight, wave action and rubbing but does not fully degrade. The smaller particles (<0.5 millimetres diameter, also known as microplastics) have been found to sink down into the water column, so that oceans are being polluted by plastic not just at the surface but at depth. Some sea creatures ingest these particles, mistaking them for plankton, and then as their stomachs are full of plastics the creatures can starve. Modelling suggests that the microplastics deeper in the water column are transported from subtropical and subpolar regions by subsurface currents towards the poles, which matches the accumulations found in sea ice and polar waters. Thus, surface currents concentrate larger, surface-floating plastic debris in gyres, while deeper currents concentrate microplastics in polar areas.

There are some organisations, such as The Ocean Cleanup, who are developing techniques to collect surface-floating macroplastics from oceans. These can make use of targeted activities in the oceanic gyres, where macroplastic concentrations are greatest, but the challenge is enormous. The macroplastics collected are recycled into other products. However, it is critical to reduce global plastic waste and to prevent it from entering oceans in the first place. This is especially necessary given that microplastics cannot readily be extracted from the oceans.

While surface currents are largely driven by wind, deep-ocean currents are driven by differences in water density. This deep-ocean circulation system is called the **thermohaline circulation** system (Figure 3.5). There are two important areas where deep-water currents form. The first is in the North Atlantic/Arctic Ocean and the second is in the Antarctic Ocean. In the far North Atlantic, salty water from the Gulf Stream moves north into the Arctic and cools. Being saline and cool it is denser than surrounding waters and so it sinks and then flows south, forming the main deep-water current of the whole Atlantic. There have been concerns that ice melt in the North Atlantic region caused by global warming might supply lots of fresh water, which would reduce the salinity of the Gulf Stream water so that it is not dense enough to sink. This would then cause the whole thermohaline circulation system to severely weaken. Research has suggested that the circulation has slowed by about 15 per cent in recent decades, although there is still some debate about when this slowing down commenced. This could result in reduced heat transfer from the equator, resulting in much colder climate conditions at higher latitudes. For example, the existing mild climate of north-west Europe could resemble the much colder climate of north-east Canada (e.g. Labrador). This would lead to further feedbacks as reflection of the Sun's energy by enhanced snow and ice cover in Europe would cool the climate further. There is some evidence that strong slowing of the thermohaline circulation has occurred naturally in the past, with important climate implications, and this is discussed in Chapter 4.

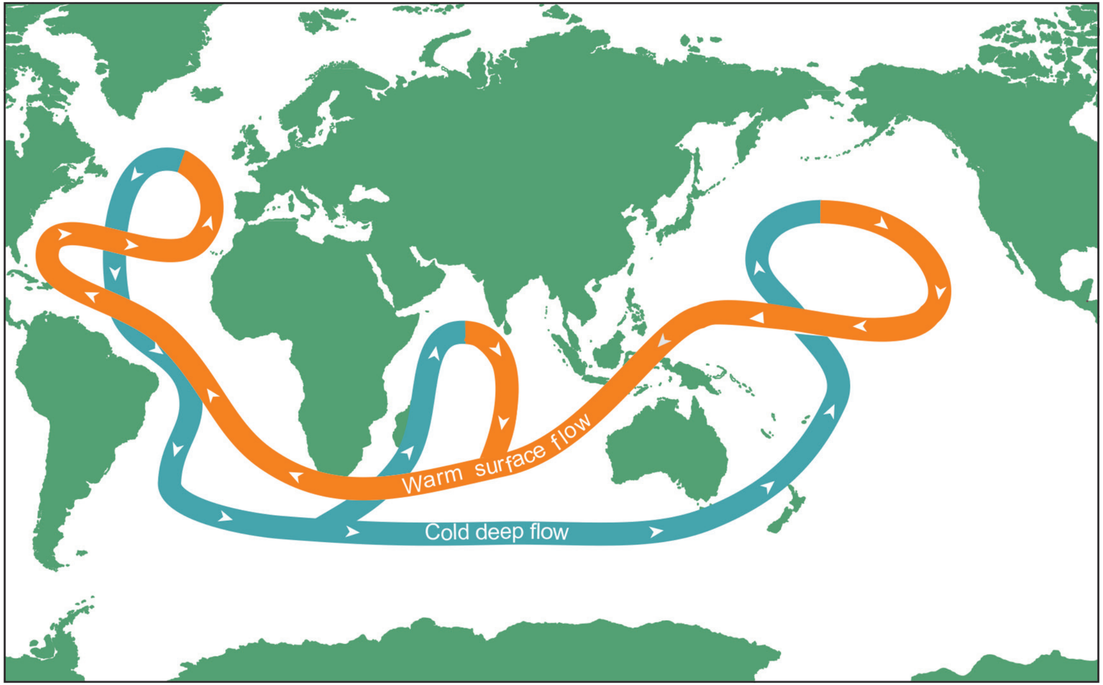

Figure 3.5 A schematic map of the thermohaline circulation system.

Sinking water has to be balanced by rising water coming to the surface. Deep water upwelling occurs thousands of kilometres away from the sinking zones. Along several eastern edges of continents where surface water is driven offshore by winds, this surface water is replaced by deep water from below. This upwelled water is often rich in nutrients that have fallen to the depths of the ocean. When these nutrients reach the surface at upwelling zones, and where there is sufficient light, the nutrients can be utilised by plankton, which in turn can maintain rich fisheries. One such example of a rich upwelling zone is off the coast of Peru.

INTERANNUAL CLIMATE VARIABILITY

While there are rich fisheries off the coast of Peru because of an important upwelling zone, these fisheries often collapse because of a specific event every few years that directly demonstrates the feedbacks between the oceans and the atmosphere in controlling the climate. This event is part of the **El Niño Southern Oscillation**. The Southern Oscillation is characterised by an exchange of air between the south-east Pacific (high pressure) and the Indonesian equatorial region (low pressure). Most of the time the trade winds are strong and converge over the warm waters of the western tropical Pacific where there is low pressure and lots of rainfall. During this time the ocean surface in the eastern tropical Pacific is relatively cold and the air above the ocean and the coastal parts of equatorial South America is cold and dry: this is the neutral state. However, every few years for 6 to 18 months the trade winds relax. As air pressure falls in the east, the warmer surface ocean water and heavy rain moves east while there is a rise in surface air pressure in the west. This event is known as El Niño and brings drought to Indonesia and heavy rainfall and floods to coastal areas in equatorial South America. The movement of the warm water to the east means that the normal upwelling off Peru is capped with warm water, preventing deep, cold water from moving up to the surface. Thus, the nutrients that support the rich fisheries are cut off.

The effects of El Niño can be seen across the planet due to the fact that the whole global climate system is linked, meaning a change in one location has knock-on effects elsewhere. For example, contrary to normal conditions most El Niño winters are warm and dry over western Canada and wet from Texas to Florida. Damage from floods and landslides caused by very high rainfall in southern California has been linked to El Niño, along with Indonesian forest fires, Australian bush fires and drought, and crop failures and famine in south-central Africa (e.g. Zambia, Zimbabwe, Mozambique and Botswana). During La Niña, many of the reverse patterns are observed, with wetter weather for Indonesia, Australia and parts of the Amazon, and drier conditions in southern United States. Some El Niño events can be more intense and last longer than others. Forecasters have been trying to establish the likely nature and severity of each El Niño event in advance so that governments, farmers, and other interested parties can prepare for the changes (e.g. by sowing different crops than normal or saving additional water before drought develops). Researchers have also been trying to establish how climate change and Southern Oscillation events might interact in the future (Box 3.2).

The El Niño Southern Oscillation is not the only example of interannual variability. For example, there is the North Atlantic Oscillation, which when in a ‘positive’ phase increases the rainfall across northern Europe (with less across southern Europe) and results in milder northern European winters, while in a negative phase leads to less rainfall in northern Europe and more in southern Europe and north Africa. The North Atlantic Oscillation can remain in a positive or negative phase for several years or even decades.

It is also worth noting that other isolated natural events can cause variability in the Earth’s climate. For example, volcanic ash from the eruption of Mount Pinatubo in 1991 in the Philippines darkened skies around the world for over a year. The dust reflected more of the Sun’s energy back into space and therefore the Earth was a little cooler that year. Following the eruption of the Tambora in 1815, unusually cold temperatures caused crop failures and famine in North America and Europe for two years.

BOX 3.2 EL NIÑO AND CLIMATE CHANGE

No two El Niño events are alike and they vary in intensity. Currently there is an extremely strong El Niño about once every 20 years. However, it is thought that with climate change, extreme El Niño and La Niña events may increase in frequency to every 10 years by the end of the twenty-first century, and the events themselves might be stronger than they are now. In addition, El Niño and La Niña events in the future may amplify other impacts of climate change. Some impacts are already more severe than previously observed, such as extensive bleaching of coral (due to the sudden warming in surface ocean temperatures) and increases in storm activity in the Pacific. It has been predicted that rising global temperatures will leave some regions, such as south-west United States, more vulnerable to severe droughts and wildfires in the future. This means that La Niña events could have **compound effects** in that region by increasing wildfire and heatwave risk even further under future climate change. During El Niño, the cooler and wetter weather in that region is likely to intensify, enhancing flood risks.

REGIONAL CLIMATE AND WEATHER

Climate is the long-term average of daily weather conditions occurring at a location. The decline in the Sun's energy received at the Earth's surface with latitude (Figure 3.1) is important for determining the climate of a particular region. However, the distribution of oceans and continents and the circulation of the oceans and atmosphere are also important. This means that two locations at the same latitude can experience very different climates and different types of weather conditions. Northern Scotland has mild winters while Labrador, at the same latitude in north east Canada, has very cold winters. The following sections outline some characteristics of climate and weather experienced in key global climate zones.

POLAR CLIMATE AND WEATHER

Polar climates consist of two main categories: ice cap and tundra. Polar ice caps are found in central Greenland and the Antarctic. They are dominated by high pressure and are extremely cold. Summer temperatures are generally below 0°C and winter temperatures below −40°C. In parts of the Antarctic the average annual temperature can be close to −50°C, with temperatures close to −90°C having been recorded there. The air is very dry and there is little precipitation. Much of the area of the polar ice caps can be officially classified as desert as there is such little precipitation, typically less than 100 millimetres per year. This low precipitation in polar regions is why many physical geographers differentiate between hot deserts and cold deserts. The ice cap environment is made even worse for survival by the strong 'katabatic winds', which occur because the ice caps cool the air around them, causing it to become denser and to sink from the high centre of the ice caps (the Antarctic interior is 3 500 metres high) towards the coast.

Polar tundra climates are found in northern Scandanavia, Siberia, Iceland, coastal Greenland and high latitudes of North America (see Figure 6.3 in Chapter 6). The temperature of the warmest month in the polar tundra is above 0°C but below 10°C. Winter temperatures are generally low (average below −25°C). Mean annual precipitation tends to be less than 300 millimetres. The weather is dominated by the prolonged winter with dry, clear, high-pressure conditions. Although summer sunshine is weak in polar tundra regions, the long daylight hours can result in melting of the snow cover for short periods, allowing the upper soil layers to thaw. The long daylight provides a short summer growing season (see Chapter 6).

MID-LATITUDE CLIMATE AND WEATHER

The mid-latitudes are dominated by weather systems that move across the Earth. Within the high level Rossby waves (described earlier) there are zones of air convergence and divergence. Upper air convergence occurs where air slows down in the Rossby wave and divergence occurs where it speeds up. If the divergence of air high up in the troposphere is greater than the convergence of air down below near the Earth's surface then there will be a fall in surface air pressure and air will rise. This divergence, along with the Coriolis force, leads to the formation of a large, circulating, rising mass of air flowing in an anticlockwise direction. Such a feature is known as a depression. If the opposite occurs then there will be a zone of high pressure at the surface with descending air flowing in a clockwise direction. Such a feature is known as an anticyclone. Descending air in anticyclones results in the air warming, which means that it is less saturated (less close to the dew point temperature) and typically means clear conditions. However, on some occasions, there can be a **temperature inversion** whereby colder, moist air is trapped below the warmer air and a layer of cloud or fog can form, especially in winter. During summer the air below the inversion is warmed sufficiently to cause the cloud or fog to dissipate.

An air mass is a regional parcel of air which has developed over an area where it has remained for a period of days and gained particular temperature and moisture characteristics. There are four basic types of air mass: tropical maritime, tropical continental, polar maritime and polar continental. Additional extremes are Arctic maritime and Antarctic continental. Continental air masses are relatively dry and maritime air masses are relatively humid.

Polar maritime air is common in both hemispheres, with source regions in the high-latitude oceans. Polar continental air occurs in the northern hemisphere, while Antarctic continental air occurs in the southern hemisphere. Tropical maritime air is common in both hemispheres, while tropical continental air is less common due to the lack of large land masses in the subtropics (e.g. North Africa is the main one). India can be a source region for tropical continental air in winter. In the summer, the high pressure over Siberia combined with the Himalayan mountain chain restricts the movement of tropical continental air northwards.

Air masses are modified by the Earth's surface. If the surface is colder than the air mass then the stability of the air will be increased. If the surface is warmer than the air mass then stability will be decreased and cloud formation and precipitation may result. For example, when Arctic maritime air moves south over the North Atlantic the sea surface is warmer than the air mass. The sea surface warms the lowest layers of the atmosphere, decreasing stability and encouraging convection and the formation of frequent showers. Weather forecasters study the source areas for air, track their movement over a particular region and predict how the air masses might be modified by local conditions and how they might interact with other air masses.

A weather front is the boundary between two masses of air of different densities (and temperature). A cold front is where a cold air mass is moving into a warm air mass, whereas a warm front is the reverse. The rising air in depressions is concentrated along the warm and cold fronts. Rising air will produce condensation of water vapour, leading to the formation of cloud and precipitation. Rising air in depressions takes place over a large area and, while there may not always be heavy precipitation, there can be substantial amounts of precipitation over a large area. Looking at vertical slices through fronts shows they tend to gently slope at a rate of 1 metre of vertical rise for every 80 to 150 metres of lateral distance. Cold fronts tend to be steeper than warm fronts and over time a cold front tends to overtake a warm front, leading to an **occluded front**. As a cold front passes there can be a sudden drop in temperature experienced at ground level (5°C fall within 30 minutes is not uncommon).

The climate zone of the mid-latitudes is often split into two regions: the first is the western edges of the continents; the second is the interiors of the continents combined with similar conditions experienced on eastern edges of continents. The winds experienced on the eastern continental edges of the mid-latitudes (e.g. eastern side of North America and Asia) have generally had a long journey across land. Hence the climate is closely related to that in the centre of the continent. Therefore, these two areas are classified under the same climate zone.

Mid-latitude western continental edge climates are mainly found in the northern hemisphere, but New Zealand, Tasmania and southern Chile are also considered to fit this zone. The western margins of continents have mild winters for their latitude because of the warm ocean currents (e.g. Gulf Stream or East Australian Current, as shown in Figure 3.4). These climates have a small range of annual temperature with precipitation distributed throughout the year and with considerable enhancement of precipitation by coastal mountain ranges (see below). Mean winter temperatures for the mid-latitude western continental edges are typically between 2 and 8°C, with mean summer maximum temperatures between 15 and 25°C. Precipitation totals generally range from 500 to 1 200 millimetres per year. These climate zones can be windy, particularly in coastal areas, and mid-latitude depressions can bring strong, damaging winds. A Mediterranean-type climate, consisting of a mild, half-year wet winter and a half-year hot, dry summer, is found in south-west South Africa, central Chile, on south-west coastlines in southern Australia, in California, as well as in the Mediterranean itself (see also Chapter 6 for a description of the typical vegetation of Mediterranean-type climate areas). Mean winter temperatures in Mediterranean climates range from 5 to 12°C, with summer daytime maximums of 25 to 30°C. Precipitation totals typically range from 400 to 750 millimetres per year, but with a summer minimum.

The mid-latitude continental interiors and their eastern margins have cold winters (mean winter temperatures around 0°C) with frequent snowfall, which often does not melt until the spring thaw, while summers are hot and humid (around 25°C). Winters become colder further north (of 45°N) or west into the centres of the mid-latitude continents and summers also become cooler and less humid. North of 50° there are severe winters and relatively short summers (maximum of three months with mean temperatures above 10°C). Precipitation tends to be distributed throughout the year in the mid-latitude continental interiors but generally with a peak in the summer mainly resulting from convective showers and weak frontal systems. Total annual precipitation is low (below 500 millimetres), but the cold winter and presence of summer precipitation provides sufficient moisture for plants (e.g. the main wheat-growing areas of North America). The regions above 50° in the mid-latitude continental interiors are dominated by high pressure in winter but frontal weather systems at other times of the year. Winter mean temperatures can be less than −25°C in the coldest month and as low as −50°C in Siberia.

TROPICAL AND SUBTROPICAL CLIMATE AND WEATHER

Close to the equator the Coriolis effect is negligible and so the weather is not dominated by the movement of large circulatory weather systems. Here, air simply flows from high to low pressure. However, flow from the north-east and south-east trade winds helps air to converge into the intertropical convergence zone (see above). Air meeting at this zone combined with the warm conditions forces upward movement and the formation of clouds. Close to the equator the weather is dominated by frequent convectional clouds and plentiful rain, with some parts of Amazonia and West

Africa receiving over 4 000 millimetres per year. Mean temperatures in the equatorial regions are around 27°C throughout the year. Equatorial climates are not the same in each place, however, as topography and proximity to the oceans play a role in altering temperature and rainfall.

Moving north or south away from the equator into the trade wind region, between about 5° and 20° latitude, a rainy (summer) and dry (winter) season becomes clearer. The trade winds produce a steady, but not particularly severe, wind regime. However, tropical depressions can form over the oceans, some of which become tropical cyclones. These cyclones are known as hurricanes in the Atlantic and typhoons in the west Pacific. Tropical cyclones require high sea surface temperatures (at least 27°C) but do not occur close to the equator as the Coriolis force is too weak. They only form over oceans because the energy to maintain their strength comes from the heat released when water vapour condenses to form clouds. When the moisture source is restricted as the cyclone hits land then the storm dies down. Tropical cyclones have high rainfall intensities and, as they travel slowly, any one place by which they pass can receive very high rainfall totals. The rainfall over two to three days from a single hurricane can be several hundred millimetres.

The climate of the trade wind belt also includes the monsoon, which occurs in Asia, west and east Africa, Australia and a weaker version over southwestern USA. The monsoons produce regions that experience an exceptionally wet rainy season. Monsoon regions are all associated with the switching of the wind direction. In winter, winds blow off the relatively cold continent towards the warm ocean (warmer air rising above the oceans draws in air to replace it from the nearby continent). There are therefore stable, dry conditions over land. However, as the land warms in the summer then the wind reverses due to the low pressure (rising air) at the surface that forms over the warmed continent. Air from over the ocean is drawn towards the land to replace the rising continental air. In doing so the oceanic air brings lots of moisture with it. These changes are also supplemented by changes in the location of the jet stream above. In Asia, the dry northerly wind over India reverses direction in May/June, and warm, humid air from the Indian Ocean flows from the south until around October, bringing torrential rains. The rains are not continuous, but in some mountainous places the amount of rainfall can be 10 000 millimetres per year. Monsoons are not always the same each year and El Niño years in Asia can be associated with the failure of the monsoon rains, sometimes leading to disastrously low crop yields.

In contrast to the wet conditions near the equator, the major deserts are commonly found around 30° latitude, coinciding with the zone of descending, dry air from the Hadley circulation cell. The driest hot deserts are found in the western coastal regions of the continents where the subtropical anticyclones are most intense. The main features of hot desert weather are a wind, which increases aridity, and high daytime temperatures (often over 35°C). The dry air and clear skies produce large daily ranges of temperature (as much as 20°C in some places) and night temperatures can even drop below freezing in places.

MOUNTAIN CLIMATE AND WEATHER

Hills and mountains can substantially modify regional and local climate. There tends to be a greater amount of the Sun's energy received in the form of ultraviolet energy in mountainous areas. This can be harmful to humans, causing skin cancer if regular, high doses are received. Pressure and temperature normally fall with increasing altitude within the troposphere. This means that typically it becomes cooler as you ascend a mountain. The higher and more isolated a mountain, the more the air temperature will resemble that of the free atmosphere rather than the atmosphere heated a short distance from the Earth's surface.

If there are light winds and clear skies at night, this can allow the ground to cool, forming a strong temperature inversion which draws air down the slopes into cooler valley bottoms. If there is a cloud cover or strong winds then this process will be restricted. During the day, if the Sun's energy is strong and the slopes are steep sided, the air can be warmed not only from below but from the sides of the steep slopes too. This leads to local winds flowing up the slopes of some valleys. It is these features that are often utilised by thrill-seekers who go hang gliding or parasailing in hilly areas.

Precipitation totals tend to increase with altitude. However, this increase is greatest in mid-latitudes. As air moving over the land surface reaches the mountain range it is forced to rise over the mountains. The cooling of the air that occurs as it rises reduces the temperature to the dew point. Further rising air leads to the formation of clouds and precipitation. Thus, in the mid-latitudes, mountains enhance precipitation, even in relatively dry areas. However, in many tropical areas the increase in precipitation with height is more complex, often due to strong convective activity which repeatedly drives rainfall upwards within clouds and due to the fact that as soon as raindrops fall out of the cloud base many will evaporate. Thus, the peak rainfall rate in tropical areas tends to be at the base of clouds around 1 000 to 1 500 metres altitude. For example, on Mauna Loa in Hawaii the rainfall is around 5 500 millimetres per year at 700 metres altitude but is only 440 millimetres at the top of the mountain at 3 300 metres altitude.

If the rising air over mountains has been forced to the dew point and cloud formation and precipitation has occurred, it means that when the air falls again, once it passes the mountain range and warms, then it will be much drier than it was before it encountered the mountain range. Therefore, in the lee (downwind) of mountains the climate can be significantly drier than on the windward side of the mountains. Since mountains are regular barriers to air flow, they are also regular locations for heavy precipitation. In fact, on the island of Kauai (part of the Hawaii chain of islands) in the Pacific, Mount Waialeale receives an amazingly high average of around 12 500 millimetres of rainfall per year on its wind-facing side. However, on the leeward side of the mountain, the slopes only receive 500 millimetres of rainfall per year. Nearby locations over the ocean where there are no mountains would normally receive around 640 millimetres of rainfall per year. The warm and dry wind that blows down lee slopes of hill and mountain ranges is called a **föhn wind**. These winds are increased further by the wave effect of fast moving air being forced over a mountain range, a little like flowing water over a pebble. There are different local names for föhn winds, such as the Chinook in North America or Zonda in Argentina. These winds can be very important because when they start to flow, they can increase local air temperature by up to 25°C in an hour, causing sudden snow melt or avalanche risk and influencing plant growth. In fact, in Canada, in the lee of the Rockies, temperature rises of over 20°C have been recorded in just a few minutes as the wind begins to flow. This can bring plants out of their winter dormancy only to be damaged when cold weather returns once the föhn wind has stopped. As the winds are warm and dry, they may even increase fire risk.

Normally, surface friction reduces wind speeds by about 30 per cent. However, exposed peaks and ridges can have higher wind speeds as there is less surface friction around them, and so the wind experienced in mountains can be different to that in the lowlands because of the nature of the terrain (and not necessarily because of the altitude). Furthermore, if wind is funnelled through gaps or valleys between individual peaks then it can be more intense at those locations. The winds in mid-latitude mountains are also influenced by the prevailing westerly winds and these winds are generally faster at higher altitudes in the troposphere. However, in the tropical and subtropical trade wind belts, the north-east and south-east trade winds generally weaken with height. Therefore, wind speeds can be low on tropical and subtropical mountains.

More precipitation may fall as snow in mountain regions, which can accumulate over time. Box 3.3 describes rain and snow formation. Often, snow-melt in the spring can produce large seasonal river flows. Where there are glaciers, these can modify the local climate. Glaciers cool air in contact with the surface and depending on the moisture content of the air they can act as either a local moisture source or sink. For example, if the air is warm then water vapour can undergo **sublimation** (change directly from the solid to gaseous state) into the air above the glacier, which uses energy and thereby cools the air.

BOX 3.3 RAIN AND SNOW FORMATION

Clouds contain tiny droplets of condensed water. The droplets are very light and remain suspended in the air. As these droplets move around, some of them collide and join together. As they grow bigger they become heavier and can start to fall, as the gravitational pull drawing them down will be greater than the force of the rising air keeping them buoyant. This process is typical in warm clouds, and the deeper the cloud the bigger the drops will grow and the faster the rainfall.

In the middle and high latitudes, however, many clouds form where it is well below 0°C. Here, clouds contain ice crystals as well as water droplets. The latter exist because the very small size of the droplets means they do not freeze straight away and they become 'supercooled' (the Bergeron process). A peculiar property related to the lower saturation point of ice compared to that of water means that some of the water droplets evaporate and then freeze onto the ice crystals, causing them to grow. As they become heavier and fall, they also collide and stick to each other, forming snow. As the flakes fall, they may warm and melt and then produce rain. Clearly, if the air is cold near the surface then this melting does not occur and snow will reach the ground.

LAND AND SEA BREEZES

Water is slow to heat up and cool down. Therefore, deep water bodies witness little daily change in surface temperature. The air over these water bodies does not experience large daily fluctuations in temperature. Over the land there are greater daily changes in air temperature, particularly in the summer when the Sun's energy is stronger and the land heats up during the day and cools down at night. Differences in local temperature across land and water result in sea and land breezes. The rising air over the land during the day draws in air sitting above the adjacent, cooler waterbody, creating a sea breeze. Sea breezes only form when there are light wind conditions during typically anticyclonic conditions. The sea breeze can exist from the surface to 2 kilometres above ground, and even up to 100 kilometres inland, and brings cooler, more humid air inland. These breezes not only develop near oceans but also develop where there are large inland bodies of water, such as the Great Lakes in North America or Lake Victoria in Africa. These breezes can bring welcome relief in the tropics, providing a fresher feel to the climate of coastal areas. When the more humid air drawn in from the nearby waterbody begins to warm over land it can start to rise, forming clouds, which is why on some summer days it can be a cloudless day inland, but when you reach the coast, it can be disappointingly cloudy. Sea breezes die off as night falls and are sometimes replaced by a weak land breeze blowing out to sea. At mid and high latitudes, during the winter half-year, the temperature of snow-covered land may remain colder than sea surface temperatures during both day and night and so sea breezes are less likely, although land breezes may form.

VEGETATION AND CLIMATE

If the land is bare then the air temperature can be affected by the darkness of the soil and the moisture content. Darker soils absorb more of the Sun's energy and therefore radiate more long-wave energy back into the atmosphere to warm it from below. Moist soils will facilitate more evaporation, thereby cooling the atmosphere as energy is used to evaporate the water.

The effects of vegetated surfaces on local climate are complex. There can be large differences from forest to grassland and from field to field depending on the crop. Different plants reflect more or less of the Sun's energy into space. Wind speeds can be affected by the height and density of vegetation. In fact, humans often plant hedgerows or rows of trees where they require some shelter from the wind. **Transpiration** from plants releases water vapour, which may cool the surrounding air.

The temperature and wind effects can be amplified in forests and it can be much cooler in forests during the day than outside of the forest. Research has also shown that forests can impact regional rainfall patterns and that continued deforestation of the Amazon at current rates, for example, might reduce regional rainfall by 8 per cent by 2050. In turn, these drier conditions may reduce hydropower capacity.

URBAN CLIMATE

Urban areas often have a different climate to that found immediately outside of the urban area. This is because of the nature of the building and road surfaces. Urban areas often have much more 'rough' and variable terrain than surrounding rural areas and this creates friction and reduces wind speed, which can also mean that air pollution (e.g. from cars) within urban areas is not dispersed very quickly. However, in strong winds, tall buildings create powerful gusts, as the wind is squeezed through the gaps between them, sometimes making it difficult to walk or open doors. Urban building materials tend to be good absorbers of the Sun's energy and do not reflect as much as non-urban land surfaces. Urban surfaces then release the absorbed energy as long-wave radiation, which warms the air from below much more than in rural areas. Furthermore, because humans heat houses, offices, shops and industrial premises, energy can be released into the urban environment, especially in the winter in mid and high latitudes. Air pollution in urban areas can also add to increased temperatures as certain aerosols readily absorb long-wave radiation emitted by urban surfaces (Box 3.4). There also tends to be less vegetation cover or open soils within urban areas and so evaporation is reduced. Evaporation requires energy to change the state of water into water vapour, thereby cooling the local environment. However, if evaporation is reduced there will be less cooling compared to surrounding rural areas. The above factors combine to create the 'urban heat island' effect, where the urban area acts as an island of warmer air surrounded by cooler rural air. The heat island is more apparent at night or when wind speeds are low. The heat island effect is related to the city size and types of building in the city. The largest heat islands are found in cities such as Beijing and New York, and a heat island effect as large as 12°C has been found in Montreal during winter.

Because of atmospheric pollution in urban areas there can be periods when the weather conditions within urban areas are a danger to human health. Exhaust fumes from vehicles within urban areas is now a widespread problem for urban air quality, especially in regions that receive high amounts of the Sun's energy, such as Athens, Mexico City and Beijing. Here, the ultraviolet radiation from the Sun reacts with vehicle emissions and produces a photochemical smog that irritates the eyes, nose and throat and creates a hazy atmosphere. It was notable that many large cities experienced less frequent smog events during the COVID-19 pandemic as fewer vehicles were on the roads and factories were shut down.

BOX 3.4 CAN URBAN HAZE WARM CITIES?

The urban heat island effect causes more intense heat waves for city dwellers. Cities are also concentrated sources of aerosols (airborne particles). The haze which results can cause respiratory problems for humans. The aerosols can also cause temperature effects. The first of these is that they block out some of the solar energy from space, reducing short-wave radiation that reaches the ground surface. However, many aerosols are also very good at absorbing and re-emitting long-wave radiation, which may increase temperatures. It has therefore been difficult to establish the net effect on urban temperatures.

A recent study of cities in China using satellite data found that there was a relationship between the amount of haze and the urban heat island effect at night time (Cao et al., 2016). While haze did not contribute to daytime warming, at night the average haze pollution contribution to temperature was around +0.7°C. Cities in semi-arid regions, such as Hami in northwest China, where aerosols often have larger particle size due to mineral dust from urban landscapes being swept up under drier conditions, were warmed much more at night by aerosols and showed larger urban heat island effects than cities in wetter areas. Thus, reducing urban haze by setting emissions policies and managing environmental processes to minimize aerosol release can benefit human health directly and reduce heat wave intensity.

SUMMARY

- The climate system is a closely integrated system involving the atmosphere, oceans, land surface and vegetation.
- The Sun heats up the Earth's surface, which in turn warms the atmosphere from below.
- Greenhouse gases trap some of this heat so that at night it does not all escape back into space.
- The energy surplus from the Sun at the equator and the deficit at the poles creates an imbalance that leads to heat transfers through air and water.
- The Earth's rotation has a large impact on air and water movements, while the Earth's tilt causes seasonal differences in the receipt and redistribution of the Sun's energy.
- There is a zone of moist, warm, ascending air near the equator which yields a plentiful rainfall supply.

- Tropical cyclones can form over warm oceans and bring large amounts of rainfall and damaging winds to coastal areas.
- At around 30° north and south there are major zones of descending air which are clear and dry, and this is where most of the world's large deserts are found.
- At the surface in middle latitudes, the main features are easterly moving cyclonic and anticyclonic systems, which bring highly variable weather conditions.
- Continental interiors have different climates to those areas close to the oceans: ocean circulation, which transports heat energy, can have a major influence on the climate of continental margins.
- Polar climates tend to be cold and dry, being associated with descending air and clear conditions, but they can be very windy.
- Topography plays an important role in modifying the local and regional climate; upland areas are often more windy and can receive more precipitation. However, the downwind side of mountain areas can be much drier than upwind.
- Sea and lake breezes can influence the local climate of areas near water bodies.
- Urban environments tend to be warmer than rural environments due to the use of building materials that absorb the Sun's energy, air pollution that traps more heat, local heat escaping from buildings that humans have heated and reduced evaporation.

FURTHER READING

Cao, C., Lee, X., Liu, S., Schultz, N., Xiao, W., Zhang, M., Zhao, L. (2016) Urban heat islands in China enhanced by haze pollution. *Nature Communications,* 7: 12509.

This journal article is of relevance to Box 3.4, providing more information about the evidence for haze effects on urban warming.

Holden, J. (ed) (2017) *An introduction to physical geography and the environment* (4th edition). Harlow: Pearson Education.

Focus on three, well-illustrated, textbook chapters by specialists covering more detail on 'Atmospheric processes' (pp. 137–174), 'Global climate and weather' (pp. 195–228) and 'Regional and local climates' (pp. 229–250).

Pinet, P.R. (2019) *Invitation to oceanography* (8th edition). Sudbury, Massachusetts: Jones and Bartlett Publishers Inc.

A classic textbook outlining all aspects of ocean processes.

Rohli, R.V. and Vega, A.J. (2017) *Climatology* (4th edition). Burlington, Massachusetts: Jones and Bartlett Publishers Inc.

Colourful textbook with a companion website.

Saunders, R. (2021) The use of satellite data in numerical weather prediction. *Weather, 76*: 95–97.

This short article provides a very useful and succinct review of the types of satellites (including a brief history), their orbits and types of measurements that are provided, and how satellite data are used in weather forecasting.

4

CLIMATE CHANGE AND THE CARBON CYCLE

THE CARBON CYCLE

Carbon is an element present in solid form in air and rocks and in dissolved form in fresh and ocean waters, and occurs in all living matter. It is the fourth most common element in the universe and is essential for life. While it only makes up about 0.025 per cent of the Earth's crust, it forms more **compounds** than all of the other elements combined, meaning that it commonly combines with other elements to make different substances. Carbon moves through and across the Earth in several forms and via a number of processes in what is known as the carbon cycle (Figure 4.1). The carbon in your body has been recycled billions of times, forming parts of lots of other organisms, rocks, water bodies and the atmosphere. In your studies you may hear about two types of carbon: 'organic carbon' and 'inorganic carbon'. The former is the carbon found in compounds that are *only* created by living things (these compounds include sugars, carbohydrates and DNA), while inorganic carbon is formed from non-living materials, such as minerals. However, these terms can sometimes be confusing because carbon dioxide, for example, is released by living things but classified as inorganic carbon – that is because it is also produced by other non-living mechanisms and not solely by living things.

The natural carbon cycle involves the transfer of carbon from the atmosphere to the land and ocean, where it remains in forms such as living matter, rock, soil or dissolved compounds before eventually returning to the atmosphere. The total amount of

DOI: 10.4324/9781003124405-4

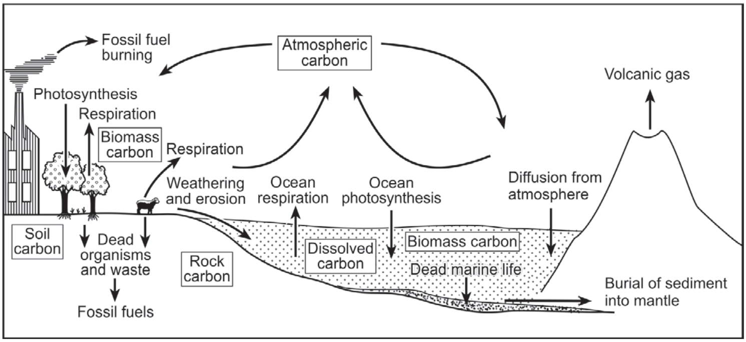

Figure 4.1 The main components of the global carbon cycle.

carbon on Earth remains the same through time but the carbon exists in different forms. The carbon cycle operates over both long (millions of years) and short (hours/days) timescales. There is much variability in how long carbon will be stored in particular locations, but on average a carbon atom will stay in the atmosphere for 5 years and in the oceans for 400 years.

One of the main mechanisms for the transfer of carbon out of the atmosphere is plant photosynthesis, whereby the Sun's energy allows plants on the land or near the surface of the oceans (such as microscopic algae) to transform carbon dioxide into carbon and oxygen. The carbon helps form the structure of the plant material while the oxygen is immediately released into the atmosphere. Most biological matter is about 40 to 60 per cent carbon when it is dried and the constituent parts are weighed. Carbon is transferred to higher level organisms when animals eat plants. Respiration uses oxygen to break down organic matter into carbon dioxide and water. Unlike photosynthesis (which only occurs in plants), respiration occurs in plants and animals. For animals, respiration is the breakdown of food by oxygen to release the energy stored by the food, and this produces carbon dioxide. Death of plants and animals means that carbon can be returned to the ground or the sea floor. Some of this is very quickly returned to the atmosphere (e.g. by fire) or decayed by bacteria, which form carbon dioxide, whereas some may be stored in the soil or ocean sediments for longer periods. A further important transfer mechanism for carbon dioxide is that which occurs between water bodies and the atmosphere when there is a concentration difference between the two. If the dissolved carbon dioxide concentration in a water body is lower than in the atmosphere then there is a **partial pressure** difference for that gas, which means that carbon dioxide will slowly diffuse into the water body. If the situation is reversed, the carbon dioxide will be released from the water body into the atmosphere (think of opening a can of soda).

Temperature is an important factor controlling the carbon cycle, the partial pressure of a water body, rates of carbon uptake and release from water bodies and carbon uptake by plants. Rates of growth and decay are much faster in tropical environments and slower in cold environments. In the oceans, photosynthetic activity occurs mainly in the upper 50 metres but varies widely depending on temperature and the amount of nutrients available in the water to support life. The detritus of dead plant and animal material that falls through the oceans towards the floor can be partly dissolved into the seawater, whereas some of the carbon can remain stored for thousands or even millions of years on the ocean floor. However, deep-ocean currents can stir up the sediment and bring some of the carbon back to the surface to later be released back to the atmosphere. The oceans currently store around 60 times more carbon than is in the atmosphere and 20 times more than in the ecosystems and soils on land.

Over timescales of millions of years, rock weathering (see Chapter 2) adds inorganic carbon to river water flowing into the oceans. Organic carbon can also be moved around the landscape by water and wind erosion, and it can be dissolved by rainwater and washed from plants and soil into water courses. Once in the ocean this carbon can be extracted by marine animals to help form their shells and bones, and when they die the carbon in the sediment can drift to the ocean floor. Some of this can be returned to the ocean surface by deep upwelling currents, but some can build up as a deep deposit. The slow movement of the Earth's plates means that ocean floors eventually get swept under the continents (see Chapter 2 for an explanation) and when this happens the sediment can be heated, melted and released back into the atmosphere (e.g. vast amounts of carbon dioxide can be released during volcanic eruptions). Additionally, carbon can get locked away within the Earth's surface for a long time when animals and plants die and the carbon enters the ground. For example, large carbon stores can build up in peatlands where the saturated environment reduces the rate of plant decay and thick organic deposits can therefore accumulate. Over long time periods this deposit can get buried and compressed and form coal or oil. Human extraction of this 'fossil carbon' for fuel then releases the carbon back into the atmosphere when it is burnt.

During combustion of fossil fuel each carbon atom is combined with two oxygen atoms to create carbon dioxide. A carbon atom has an atomic weight of 12, while a carbon dioxide molecule has the atomic weight of 44, which is 3.67 times that of carbon. Therefore, you will often see reported values of one of: i) carbon release; ii) carbon dioxide release (which will be 3.67 times the mass of carbon released); and iii) carbon dioxide equivalents (sometimes written as CO_2-eq). The latter is used when different greenhouse gases (see Chapter 3) are being discussed in terms of their global-warming potential, so all are put into the equivalent amount of carbon dioxide for ease of comparison. That is because for a given amount of gas, one compound can have a much stronger global-warming potential than another. For example, methane is 28 times more powerful in terms of global-warming potential over a 100-year period than the same amount of carbon dioxide. Greenhouse gases that do not include carbon can also be evaluated on the same global-warming potential scale. Nitrous oxide, for example, is 265 times more powerful than carbon dioxide over a 100-year timescale.

Currently, deforestation in the tropics releases 1.5 giga (thousand million) tonnes of carbon per year. Our burning of fossil fuels releases approximately 10 giga tonnes of carbon per year. However, atmospheric carbon has been increasing by a smaller figure of around 4 giga tonnes per year. Around one third of the additional carbon is taken back up on land by plants that can grow faster with higher carbon dioxide concentrations, while another third is taken up by the oceans. These areas of net uptake of carbon are known as carbon sinks. However, these additional sinks for carbon might not continue to absorb such a high proportion of the carbon we have been releasing into the atmosphere for much longer. There is probably a threshold maximum amount of additional carbon that the oceans and vegetation can take up, and once this buffer has been used up, the rate of increase in atmospheric carbon will climb even further. Furthermore, soaking up excess carbon in the oceans causes acidification of the oceans, which can lead to mollusk shells or coral becoming more soluble, and hence ecosystems may become damaged. Worryingly, recent research has suggested that the Arctic Ocean, once thought to be a likely more important carbon sink in the future as sea ice retreats, may actually be rapidly losing its carbon sink function, perhaps due to the influx of freshwater to the system (Woosley and Millero, 2020). Box 4.1 provides evidence about the potential diminishing role of the Amazon rainforest in soaking up excess carbon.

BOX 4.1 DECREASING TROPICAL FOREST CARBON SINK

Tropical forests accounted for about half of the land-based carbon uptake and removed about 15 per cent of human-caused carbon dioxide emissions in the 1990s and 2000s. Many vegetation models previously suggested that this may continue long into the future, meaning that forests like the Amazon can keep helping to reduce the impacts of fossil fuel emissions. Such forests have undergone increasing tree growth, in line with rising temperatures and atmospheric carbon dioxide concentrations. However, research based on networks of long-term forest sampling plots in tropical Africa and Amazonia, published by Hubau et al. (2020), suggested that while both areas show these tree growth effects, they are now becoming much weaker carbon sinks than in the past. This decline in performance is because of enhanced tree mortality. Essentially, trees can grow faster (hence a period of enhanced carbon uptake) but they also die more quickly, so after a period of peak carbon uptake, that rate of net uptake will decline. Worryingly, trends suggest that by 2030 the carbon sink in African tropical forests will decline by 14 per cent from the measured 2010–15 average. It is predicted that the Amazon sink will continue to rapidly decline, reaching zero by 2035. This means that a vital net sink of human carbon emissions will disappear and there will be even stronger potential for accelerated increases in atmospheric carbon.

It is challenging to determine the exact nature of all of the feedbacks between the Earth's climate system and the carbon cycle. Photosynthesis is more efficient when there are increasing atmospheric concentrations of carbon dioxide, and this in turn will promote plant growth and take up more carbon dioxide (a negative feedback). Warmer temperatures may also enhance the growing season in some locations. Additionally, the oceans can take up more carbon dioxide when there is more in the atmosphere. However, in warmer water carbon dioxide is less soluble, and so warmer oceans may absorb less. Warming may also increase soil respiration, drought occurrence and forest dieback, which may release more carbon to the atmosphere. To understand the various feedbacks and the carbon balance, researchers monitor atmospheric carbon dioxide concentrations in different regions, use ocean buoys and ship measurements to determine the carbon dioxide concentrations in the surface oceans, measure fluxes of carbon between the land and atmosphere, or water bodies and atmosphere, and use automated eddy covariance flux towers or manually sampled small chambers. Researchers also conduct experiments in chambers where the air is enriched with carbon dioxide to see how plants and soils respond, and conduct warming studies. However, there are still major challenges to combine all of these datasets and there is a lack of flux data for most parts of the Earth system. There is, therefore, an urgent need for more investment in measurements and models to help us understand the interactions of the changing carbon cycle and the Earth's climate system.

CLIMATE CHANGE

LONG-TERM CLIMATE CHANGE

The Earth's climate has changed throughout its 4.6-billion-year history, but usually on long, slow timescales. It may be a surprise, given all of the current attention on global warming, to note that today we live in an ice age. This is evident by the fact that there are still large glaciers and ice sheets covering parts of the planet. These ice sheets have been present over the past 2.4 million years and have expanded and contracted in cycles. This period of time is known as the **Quaternary**. Before the Quaternary, the Earth was much warmer and lacked ice sheets. In fact, it was warm for around 280 million years. Further back in time there were four other ice age periods in Earth's history, which lasted for between 30 and 300 million years and were separated by very long warm periods lasting hundreds of millions of years. So, while our current Quaternary ice age sounds like it has lasted a long time, it is actually very short compared to the timescales of Earth's history. Understanding past changes to the Earth's climate provides information that is useful to our understanding of our current climate, current landscapes on Earth, and what might happen in both the near and long-term future.

The start of the Quaternary period seems to have been initiated by **plate tectonics** (see Chapter 2). The movement of the continents created the suitable conditions for the beginning of the ice ages by positioning Antarctica over the South Pole, thereby allowing a large land mass to cool and build up an ice sheet, which then further cooled the climate by becoming a large reflective body (with higher **albedo**) of the Sun's energy. The northern hemisphere continents were also huddled around the Arctic Ocean and the ocean circulation systems were fixed into new positions. These continental and ocean configurations then allowed other climate forcing factors to become important.

During the Quaternary the climate has cooled and warmed many times, and this has been associated with major advances and retreats of ice sheets. Ice advances and retreats have shaped the land surface, carving new features out of rock, depositing sediments and landforms, and therefore leaving evidence of past climates behind. When ice sheets grow, sea levels fall because water is taken out of the oceans and locked up on land. The last major ice advance peaked at around 18 000 years ago (with sea levels some 120 metres lower than at present). Only 9 000 years ago, thousands of years after the ice had retreated, it was possible to walk from Britain to Europe across land because the southern North Sea (Doggerland) and the English Channel were dry, as sea levels were still relatively low.

The last 11 700 years have been relatively warm and this warm period within the Quaternary is known as the **Holocene**. The cold periods when ice advances are known as **glacials** and the warmer periods in between are known as **interglacials**. The

Holocene is our present interglacial. Several processes have been proposed to explain natural Quaternary climate changes and many of these involve internal feedback mechanisms within the Earth's climate system. However, it has been shown that changes in the receipt of the Sun's energy on Earth help to partly explain the glacial cycles that have occurred during the Quaternary. This is known as orbital forcing or the 'Milankovitch theory', which is named after an early twentieth-century mathematician who examined the Earth's orbit around the Sun.

Orbital forcing theory is based on the idea that the amount of energy reaching different parts of the Earth from the Sun varies through time, but in a regular and predictable way. The forcing varies with three factors, as shown in Figure 4.2. First, the shape of Earth's orbit around the Sun (its eccentricity) varies over around 100 000-year cycles as it moves from being more circular to more elliptical and back again. The eccentricity effect causes the seasons

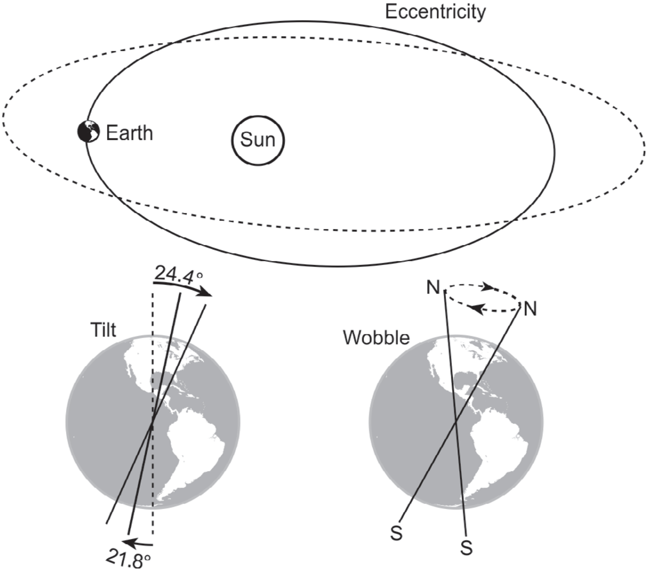

Figure 4.2 The three slowly changing and cyclical mechanisms related to the Earth's orbit around the Sun that impact the energy that reaches the Earth's surface.

in one hemisphere to become more intense while the seasons in the other are moderated. Second, the Earth's axis around which it rotates is currently tilted at 23.5°. This tilt varies through time from 21.8° to 24.4° over a 41 000-year cycle. The greater the tilt effect, the more intense the seasons in both hemispheres become; summers get hotter and winters colder. Third, the Earth has a slow wobble on its spinning axis caused by the gravitational pull exerted by the Sun and the Moon. This happens over two cycles of 19 000 and 23 000 years. The wobble determines where in the orbit the seasons occur, and most importantly the season when the Earth is closest to the Sun. The Earth currently reaches its furthest point from the Sun during the southern hemisphere winter. Therefore, southern hemisphere winters are slightly colder than northern hemisphere winters, while southern hemisphere summers are slightly warmer.

It is remarkable that scientific evidence from marine sediments and other sources shows that the timescale and frequency of the advance and retreat of ice sheets produces a close match to cycles of solar energy predicted by orbital forcing theory. For example, there have been eight large glacial build-ups over the past 800 000 years, on an approximately 100 000-year cycle, each coinciding with minimum eccentricity (Figure 4.3). Smaller decreases or surges in ice volume have come at intervals of approximately

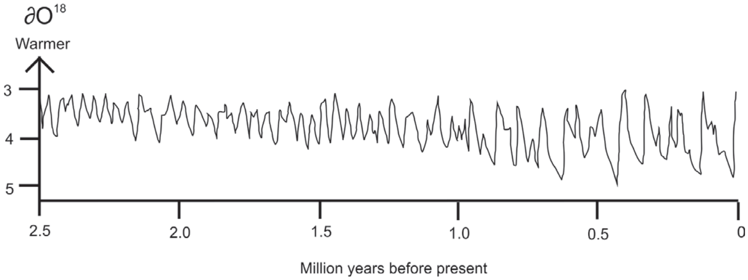

Figure 4.3 The temperature record for the last 2.5 million years, reconstructed based on oxygen isotope records in marine sediments. The isotopic record indicates how much freshwater was locked up as ice. Lighter oxygen isotopes in water are preferentially evaporated from the oceans and if proportionally less lighter isotopes return to the ocean as rainfall or riverflow, then that water is being stored in ice masses.

23 000 years and 41 000 years, in keeping with the frequency of the other two orbital mechanisms.

While orbital forcing theory matches the timing of the cool and warm periods, it is not sufficient on its own to explain the magnitude of changes in temperature experienced. In fact, there appears to be a 4 to 6°C difference between what should have occurred, in theory, due to changes in energy receipt from the Sun and the colder conditions that were manifest during the Quaternary cycles. Furthermore, the orbital cycles mathematically predict a smooth rise and fall of temperature. However, the evidence for the actual build-up and retreat of ice shows a sawtooth pattern (Figure 4.3). Over tens of thousands of years, ice sheets slowly built-up several kilometres thick, eroding the landscape as far south as central Europe and Midwestern USA. However, each cycle ended abruptly. Within a few thousand years the ice sheets melted back to present-day patterns. Hence, orbital forcing seems to be a driver of climate change through the Quaternary, but other feedbacks must be operating to account for the temperature shortfall and the sawtooth pattern. Understanding these feedbacks might be important in helping our understanding of contemporary or future climate change.

A positive climate feedback is derived from albedo, with the reflection of more of the Sun's energy from the ice causing an additional drop in temperatures, allowing ice sheets to expand further. Ice sheet expansion would have been enhanced by a fall in sea level, as ice sheets could extend further on land that was exposed, and therefore enhance albedo. Nevertheless, these feedbacks are still not sufficient to explain the magnitude of temperature changes observed.

Ice cores from ice sheets in Greenland and Antarctica contain bubbles of gas. These bubbles contain air from the time when the snow fell that later formed the ice. Data from the gas bubbles in cores near the centre of the old ice sheets in Greenland and Antarctica indicates that carbon dioxide concentrations in the atmosphere were greater during interglacials and lower during glacials. As a greenhouse gas, the changes in carbon dioxide concentrations in the atmosphere will have significantly affected the Earth's temperature. So, the key question is, what could have caused changes in atmospheric carbon dioxide concentrations during the Quaternary glacials and interglacials? One of the key theories revolves around changes to the thermohaline deep-water circulation system in the oceans (described in Chapter 3). The thermohaline circulation is driven by temperature and salt concentration gradients and acts as a strong, deep current that pumps carbon dioxide and nutrients from the surface of the oceans to the deeper waters and then returns them to the surface again. There are sensitive zones where such downwelling and upwelling occurs. If the thermohaline circulation acts as in its current fashion, then the carbon dioxide on the ocean floors would be stirred up and taken back to the surface. Deep carbon stores would not be returned to the surface as quickly if the thermohaline circulation was to slow down. The surface of the oceans would therefore be depleted of carbon dioxide and less would be returned to the atmosphere. Overall, this process would result in declining atmospheric carbon dioxide concentrations, as the ocean plants continue to photosynthesise. The energy transfer rates between the equator and poles would also be affected by changes in the ocean circulation system.

There is evidence that the thermohaline circulation system slows down during glacials. Wind enhances ocean evaporation, which makes the water more saline, making it denser, and so it sinks. However, a change in air currents around ice sheets that have formed several kilometres thick may alter wind patterns in key areas of the ocean. Reduced evaporation in sensitive areas of deep-water formation in the North Atlantic could slow down the rate of downwelling or switch it off, meaning the whole deep-ocean circulation system becomes more sluggish. Return of carbon dioxide from the deep to the atmosphere would slow, since the upwelling that occurs elsewhere would be reduced. The above theory is not entirely accepted, but it does provide one current line of thinking and it is widely believed that the North Atlantic is a very sensitive part of the Earth's climate system, with changes there potentially having global impacts. Once increased solar energy starts to warm the planet as we move out of a glacial period, then positive feedbacks could result in the rapid warming indicated by the sawtooth pattern in the evidence (Figure 4.3). This may be related to lower albedo as ice retreats and the sudden switching on of the thermohaline circulation system.

As well as the long-term cycles during the Quaternary there have also been shorter cycles, some of which are just a few decades long. Many of these fast-changing periods, however, seemed to have occurred during glacials rather than interglacials. This was probably related to interactions between ice sheet dynamics, ocean circulation and biological productivity. So far, we have only found evidence for a few rapid changes in climate during interglacial periods and it should be noted that we currently live in an interglacial. Therefore, the rapid climate changes that are occurring at the present time are highly unusual.

CONTEMPORARY CLIMATE CHANGE

Many researchers suggest that humans have modified the environment so much that we now live in a human-dominated climate-environment system known as the Anthropocene (see Chapter 1). While the identification of this geological epoch has not yet been formally ratified by international agencies, evidence for human influence on climate is compelling and clear. Over the last century, glaciers and ice caps have been receding, snow cover has reduced and sea levels have risen by about 20 centimetres (Figure 4.4b). The globally averaged and combined land and ocean surface temperature has warmed by about 0.85°C since 1880 (Figure 4.4a). The concentrations of greenhouse gases, such as carbon dioxide and methane, in the atmosphere are greater than at any time in at least the last 800 000 years, and their concentrations have risen sharply in recent decades (Figure 4.4c). Global average atmospheric carbon dioxide concentrations in 2021 were around 415 parts per million, compared with only 280 parts per million before the Industrial Revolution and around 340 parts per million in 1980.

The Intergovernmental Panel on Climate Change (IPCC) is the key body that synthesises all of the latest information on contemporary climate change. They will report a synthesis of their latest findings (AR6) in mid 2022, but their 2014 report (AR5) stated: ‘human influence on the climate system is clear, and recent anthropogenic emissions of greenhouse gases are the highest in history. Recent climate changes have had widespread impacts on human and natural systems.’ They also noted ‘warming of the climate system is unequivocal, and since the 1950s, many of the observed changes are unprecedented over decades to millennia’ (IPCC, 2014).

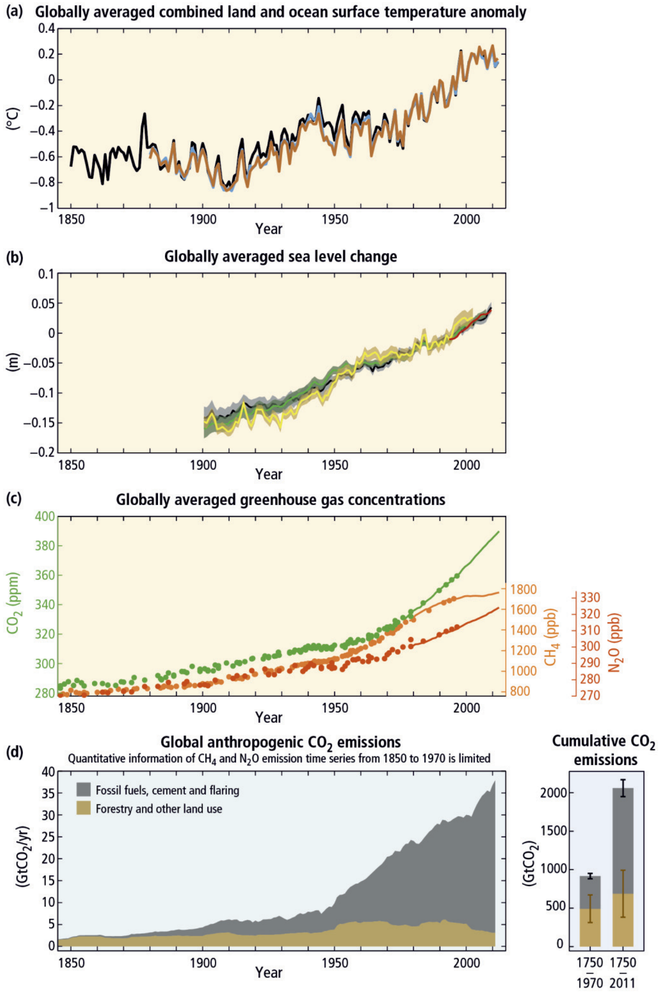

Figure 4.4 Observations and other indicators of a changing global climate system. Observations: (a) Annually and globally averaged combined land and ocean surface temperature anomalies relative to the average over the period 1986 to 2005. Colours indicate different data sets. (b) Annually and globally averaged sea-level change relative to the average over the period 1986 to 2005 in the longest-running dataset. Colours indicate different data sets. All datasets are aligned to have the same value in 1993, the first year of satellite altimetry data (red). Where assessed, uncertainties are indicated by coloured shading. (c) Atmospheric

concentrations of the greenhouse gases carbon dioxide (CO_2, green), methane (CH_4, orange) and nitrous oxide (N_2O, red) determined from ice core data (dots) and from direct atmospheric measurements (lines). Indicators: (d) Global anthropogenic CO_2 emissions from forestry and other land use as well as from the burning of fossil fuels, cement production and flaring. Cumulative emissions of CO_2 from these sources and their uncertainties are shown as bars and whiskers, respectively, on the right-hand side. The global effects of the accumulation of CH_4 and N_2O emissions are shown in panel c.

(Source: IPCC, 2014, Figure SPM.1.)

The recent temperature increase has been widespread over the globe but more exaggerated at higher northern latitudes. For the Arctic, the temperatures over the last 100 years have increased at twice the rate of the global average. Northern hemisphere spring snow has decreased in extent since the 1990s and average Arctic sea-ice cover has decreased by about 4 per cent per decade since 1979. Numerous other long-term changes in climate have been observed. For example, there have been increases in annual rainfall since 1901 in the northern mid latitudes, and the datasets provide a high confidence of these increases since 1951 (Figure 4.5). There has been an increase in the number of heavy precipitation events in a number of regions. Changes in many extreme weather and climate events have been observed since around 1950. The IPCC consider that it is *likely* that the frequency of heat waves has increased in large parts of Europe, Asia and Australia, and that human influence has more than doubled the probability of occurrence of heat waves in some locations. It is *very likely* that human influence has contributed to the observed global-scale changes in the frequency and intensity of daily temperature extremes since the mid-twentieth century. The upper 75 metres of ocean waters has warmed by 0.11°C per decade since 1971, with warming of deeper layers measured since the 1990s.

Some have argued that the above changes are naturally driven. For instance, solar cycles affect the energy received by the Earth. Volcanic eruptions eject greenhouse gases into the atmosphere but these have not become more frequent. The IPCC estimated in AR5 that only about 2 per cent of radiative climate forcing since 1750 can be explained by natural causes while the rest is anthropogenically derived. They noted that greenhouse gas emissions

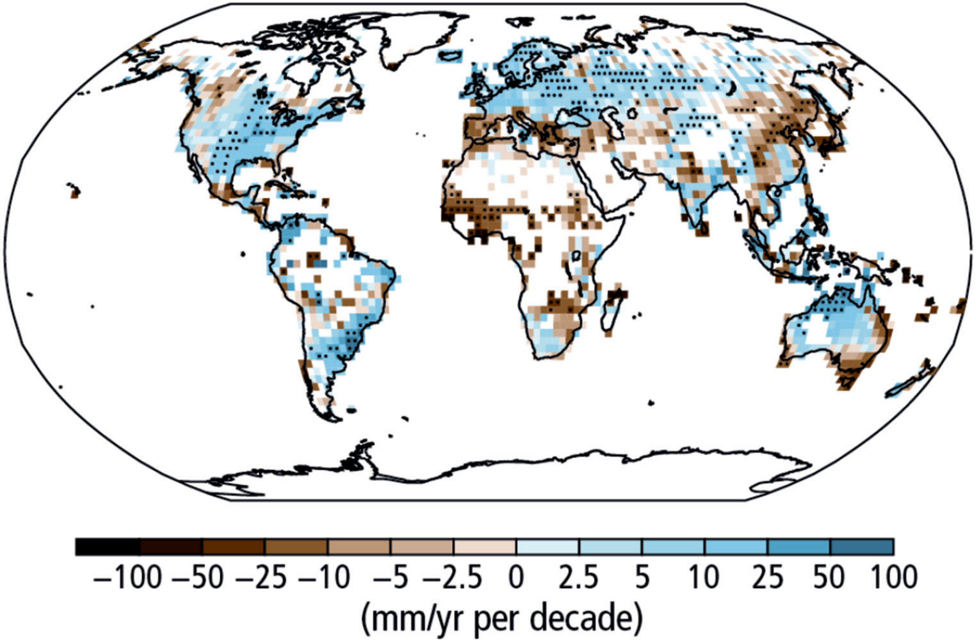

Figure 4.5 Map of observed precipitation change from 1951 to 2010. Trends have been calculated where data availability permitted a robust estimate (i.e., only for grid boxes with greater than 70 per cent complete records and more than 20 per cent data availability in the first and last 10 per cent of the time period), other areas are white. Grid boxes where the trend is significant, at the 10 per cent level, are indicated by a + sign.

(Source: IPCC, 2014, Figure 1.1.)

from human activities 'are *extremely likely* to have been the dominant cause of the observed warming since the mid-20th century'. The IPCC have also shown that Global Climate Models, which simulate the production of greenhouse gases, show the same pattern of global warming as we have observed. When these models are run without adding human production of greenhouse gases, the models show no appreciable global warming (see http://www.ipcc.ch).

Figure 4.6 shows the key sources of greenhouse emissions that are the very dominant drivers of contemporary climate change. Energy generation for electricity, heating and transport accounts for almost three quarters of the total. Agriculture and land use are key emissions sources, accounting for 18 per cent of the total. Methane released from rice production and ruminant livestock

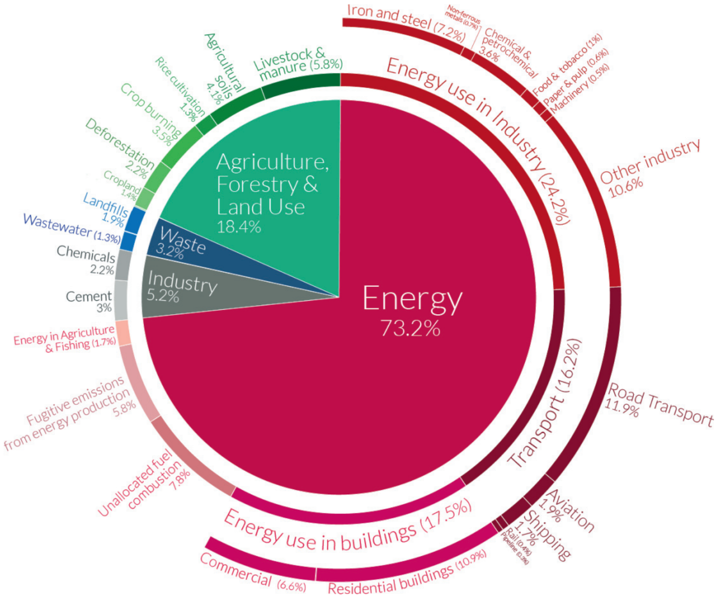

Figure 4.6 The proportion of global greenhouse gas emissions by sector, using CO_2-eq.

(Source: Hannah Ritchie (2020) Our World in Data, https://ourworldindata.org/ghg-emissions-by-sector#note-2. Creative Commons CC BY 4.0: https://creativecommons.org/licenses/by/4.0.)

production is important along with nitrous oxide released following the application to soils of nitrogen fertilisers for crops. Carbon dioxide release from cement manufacture alone accounts for 3 per cent of global emissions. Humans have also emitted fumes and smoke from industrial processes and transport. These **aerosols** partly shield the planet from the Sun's energy, scattering and reflecting it. However, not all aerosols behave in the same way. Aerosols from vegetation burning, 'black carbon' and the 'brown clouds' that come from some industrial or urban sources may have the opposite effect. The former seem to cause warming due to absorption of solar radiation, while the latter can result in some cooling due to reflection of solar radiation back to space.

PREDICTIONS OF FUTURE CLIMATE CHANGE

Predicting how the climate will behave in the near future is an important research area as it can provide vital information for policy makers and planners. It is undoubtedly the case that continued emission of greenhouse gases will cause more warming and major changes in all components of the climate system, increasing the likelihood of severe and irreversible impacts for people and ecosystems.

In order to make clear the different future choices, the IPCC use Representative Concentration Pathways (RCPs) to make projections based on population size, economic activity, lifestyle, energy use, land use patterns, technology and climate policy. The RCPs include a stringent mitigation scenario (RCP2.6), two intermediate scenarios (RCP4.5 and RCP6.0) and one scenario with very high greenhouse gas (GHG) emissions (RCP8.5). RCP2.6 is representative of a scenario that aims to keep global warming below 2° C above pre-Industrial temperatures. Surface temperature is projected to rise through the twenty-first century under all assessed IPCC emission scenarios.

Future climate will depend on past emissions, which lead to warming that we are already committed to, as well as future anthropogenic emissions and natural climate variability. The IPCC special report on impacts of 1.5°C warming, published in 2018, predicted that global warming is *likely* to reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate. The

IPCC suggest that it is 'virtually certain' that there will be more frequent hot temperature extremes and fewer cold ones over most land areas on daily and seasonal timescales. It is *very likely* that heat waves will occur with a higher frequency and longer duration. The IPCC predict that global mean surface temperature change for the period 2016–2035 relative to 1986–2005 is similar for the four RCPs and will *likely* be in the range of 0.3°C to 0.7°C. This assumes that there will be no major volcanic eruptions or unexpected changes in total radiation output by the Sun. However, by the mid twenty-first century, the magnitude of projected climate change is substantially affected by the choice of emissions scenario. Compared to 1850–1900, global surface temperature change for 2081–2100 is *likely* to exceed 1.5°C for RCP4.5, RCP6.0 and RCP8.5. Warming is *likely* to exceed 2°C for RCP6.0 and RCP8.5. Modelling suggests that warming is *more likely than not* to exceed 2°C for RCP4.5, but *unlikely* to exceed 2°C for RCP2.6. It is thought that the Arctic region will continue to warm more rapidly than the global average. Changes in precipitation will vary by location, with high latitudes and the equatorial Pacific *likely* to experience an increase in annual mean precipitation. In many mid-latitude and subtropical dry regions, mean precipitation will decrease, while in many mid-latitude wet regions, mean precipitation will increase. Extreme precipitation events over most of the mid-latitude land masses and over wet tropical regions will become more intense and more frequent.

IMPACTS OF CLIMATE CHANGE

There is already solid evidence for observed impacts of climate change, which is strongly impacting natural systems. Figure 4.7 shows impacts already being felt globally where scientific evidence has been able to attribute climate change as the cause. Changes in rainfall, snow and ice is affecting the hydrology in some regions. For example, satellite images between 1984 and 2018 showed that 56 per cent of rivers globally have seasonal river ice, but that there was a 2.5 per cent decline in river ice occurrence during this period. Some species (e.g. several species of spiders, butterflies and grasshoppers) have shifted their ranges, migration patterns and interactions in response to climate change.

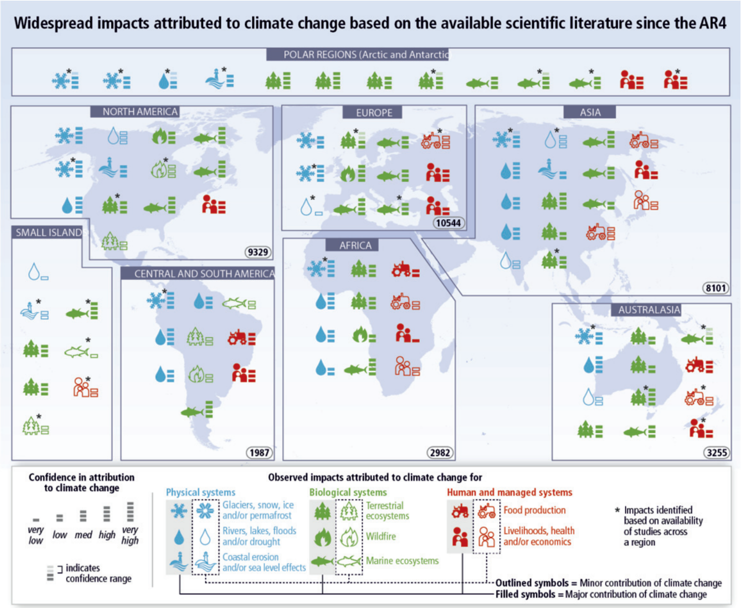

Figure 4.7 Based on the available scientific literature since the IPCC Fourth Assessment Report (AR4), there are substantially more impacts in recent decades now attributed to climate change. Attribution requires defined scientific evidence on the role of climate change. Absence from the map of additional impacts attributed to climate change does not imply that such impacts have not occurred. The number of publications supporting attributed impacts (ovals) reflect a growing knowledge base, but publications are still limited for many regions, systems and processes, highlighting gaps in data and studies. Symbols indicate categories of attributed impacts, the relative contribution of climate change (major or minor) to the observed impact and confidence in attribution.

(Source: IPCC, 2014, Figure SPM.4.)

Projecting forward, there are a wide range of potential future impacts of climate change. These may be compounded by enhanced exposure of larger human populations in zones susceptible to risks such as coastal flooding (e.g. major coastal megacities such as Kolkata, Shanghai, Ningbo, Tokyo and New York). Some examples include:

- There may be more severe weather events (storms, floods, hurricanes, heatwaves, droughts etc) as a warmer Earth means there is more energy available for atmospheric and oceanic transfers. For example, there could be a 10 per cent increase in average hurricane wind speed.
- Warming may induce sudden shifts in regional weather patterns, such as the monsoons or the El Niño Southern Oscillation (see Chapter 3). This could have major consequences for water availability, crops, flooding and wildfire in tropical regions. Some climate models predict a more frequent and intense El Niño, perhaps even causing replacement of the rainforest of Brazil with savanna.
- Melting glaciers will at first increase flood risk and then, once depleted, reduce water supplies. Large parts of the Indian subcontinent, China and the Andes rely on summer glacial meltwater to provide water for crops and to drink. Without this regular supply new ways will have to be found to capture water for around 1 billion people.
- 40 per cent of the world's population live within 100 kilometres of the coast, and two thirds of the world's largest cities are found there. Rising sea levels from ice melt and the expansion of the water upon warming could result in hundreds of millions more people enduring floods each year. More than a fifth of Bangladesh could be submerged by 2100. Sea-level rise will continue for centuries, even if the global mean temperature is stabilised. Small island states, such as the Maldives, are extremely vulnerable to coastal erosion and submergence.
- Some cold regions, such as parts of Canada, Russia and Scandinavia, may enjoy an increase in agricultural productivity as a result of warming, at least in the short term, while many of the world's main food-producing regions may become too hot and dry for crops to grow (e.g. central and southern Europe and the Great Plains of North America). Declining crop yields, especially in Africa, could leave hundreds of millions without the ability to produce or purchase sufficient food. In rice growing regions, warmer temperatures will enhance (toxic) arsenic uptake from soils to rice plants – a major problem for rice-dominant food systems.

- A nearly ice-free Arctic Ocean in the summer before 2050 is *likely* for RCP8.5. This may open routes for shipping but also potentially lead to environmental damage from mineral exploration in the region.
- The area burnt by wildfires may increase substantially. It has been estimated that the area burnt by wildfire in western United States may increase by 200 to 400 per cent for every degree of warming. These wildfires have knock-on impacts on air quality, which can be extremely detrimental to human health. They can also accelerate soil erosion, which in turn influences river water quality and aquatic ecosystems. Wildfires can also trigger long-term thawing of permafrost in high latitudes – a significant store of carbon.
- Deaths from malnutrition and heat stress may increase. Diseases are likely to spread from the tropics to the mid-latitudes as the climate warms, and outbreaks of pests may become more extreme because natural biological controls may be lacking in these newly colonised pest areas.
- Plant and animal extinction rates will increase as many species cannot migrate fast enough to keep up with the temperature changes. Most small mammals and freshwater molluscs will not be able to keep up at the rates projected under RCP4.5 and above in flat landscapes by 2100. By 2050, up to 37 per cent of species may be heading towards extinction.
- Ocean acidification will further reduce the ability of species such as molluscs, crustaceans, coccoliths and corals from forming the calcium carbonate needed for their structures, which would have knock-on effects for the food chain.
- Global economic output could be reduced by 3 to 10 per cent, with the poorest countries most badly affected.

CLIMATE CHANGE MITIGATION

It is clear from the above examples of the potential impacts of climate change that we should make every effort to reduce future climate change. Limiting climate change would require substantial and sustained reductions in greenhouse gas emissions, which, together with adaptation (see below), can limit climate change risks. There has therefore been a drive in several countries towards achieving 'net zero'. Net zero is achieved when the amount of greenhouse gas we add to the atmosphere is balanced by the amount taken out of the atmosphere each year. While action now will cost a lot, it is thought this could be a lot less costly than if we did not take action at all, given the nature of the impacts described above.

Greenhouse gas emissions can be reduced by improving the efficiency of global energy supply; making homes and industrial processes more energy efficient; enhancing technology to utilise renewable energy, such as solar, wind and tidal power; reducing demand for goods and services that emit a lot of greenhouse gases in their production; avoiding deforestation, peatland drainage (see Box 4.2) and other land management activities that increase emissions; and using lower-carbon techniques for heat generation and transport. Adopting land management strategies that capture more carbon could be important too. Tree planting is often advocated. A tree photosynthesises carbon from the atmosphere and stores it in its biomass during its lifetime. The surrounding soils also benefit and can capture carbon over even longer periods. However, researchers have found that in some places tree planting may result in negative climate impacts that outweigh the carbon capture. For example, in some high latitudes the low albedo of trees absorbs more of the Sun's energy than the surrounding landscape, thereby leading to warming. In general, however, reforestation of areas previously deforested is advantageous, and encouraging forest regrowth in tropical regions is very important.

Net zero can only be fully realised if the international community works together. There have been several attempts to persuade the international community to sign up to legally-binding climate agreements. The Paris Agreement of 2015 was a major milestone in this regard in that 196 countries signed up to limit global warming to well below 2°C, but preferably to 1.5°C, compared to pre-Industrial levels. In order to deliver such a target, countries are immediately attempting to reduce greenhouse gas emissions. However, the fickle nature of individual country leaders can ride roughshod over such agreements, as illustrated by President Trump's withdrawal from the agreement when he took office in 2017, and the reinstatement of the USA as a signatory in 2021 when President Biden took office. China, responsible for 28 per cent of the world's greenhouse gas emissions, has pledged to achieve net zero by 2060, while several other countries have now enshrined net zero in their own national laws (e.g. Sweden by 2045 and the UK, Japan and France by 2050).

Many businesses and individuals are also pledging to achieve net zero. Progress towards those pledges is often measured by calculating the **carbon footprint** of organisations and then either changing their activities or paying for carbon offsetting schemes. Thus, there is a carbon market where businesses or individuals can pay into schemes (e.g. tree planting) to offset their own emissions. However, it is sometimes difficult to establish how much 'embedded carbon' was used in making or transporting a product or in delivering a service. Furthermore, it is not clear how you would know whether or not you have already paid for that carbon through the price you paid for the product (i.e. have the company already paid a carbon tax for the carbon they used in making the product and, therefore, passed that cost on to customers). More clarity is needed on global carbon pricing so that organisations and individuals have more long-term confidence and understanding of the carbon payment system.

BOX 4.2 PEATLAND MANAGEMENT TO MITIGATE CLIMATE CHANGE

Peatlands store around one third of the world's soil carbon. They are found in tropical regions, mid latitudes and high latitudes where there is either poor drainage or plentiful rainfall to keep the system wet. Waterlogging reduces the rate of decomposition of plant material when plants die. Thus, the rate of accumulation of dead plant litter can exceed decomposition and therefore an organic-rich deposit can grow over time. Often the rates of accumulation are very slow, so it may take 1000 years for 50–100 centimetres of peat accumulation, but over thousands of years peatlands have accumulated to several metres depth in many locations. When we weigh peat after the water has been removed, we find that around half of the mass is carbon. Overall, across the globe, peatlands are net carbon sinks. However, degradation of peatlands has occurred due to their extraction for fuel and horticulture, drainage to support agriculture, forestry operations, overgrazing and atmospheric pollution. The result of degradation is that more air can get into the upper peat layers and the peat starts to decompose more quickly than it accumulates. Thus, many damaged peatlands are releasing their carbon at much faster rates than they are accumulating carbon.

Research published in *Nature* by Evans et al. (2021), based on measuring gas fluxes and hydrology at many peatland sites, showed that the depth to the saturated layer (the **water table**) was the most dominant factor controlling the greenhouse emissions from peatlands. In fact, it was such a sensitive and important control that for every 10 centimetre reduction in water-table depth, the net warming impact of carbon dioxide plus methane emissions (based on 100-year Global Warming Potentials) are reduced by at least 3 t CO_2-eq ha^{-1} yr^{-1}, until average annual water tables are < 30 centimetres. Raising water tables above that level continues to have a net cooling impact until average annual water tables are < 10 centimetres. For peatland sites where widespread ponding of water above the surface is encouraged, then they may have net warming impacts due to enhanced methane release. For peatlands where average annual water tables are just below the surface, however, they have a neutral or negative global warming potential and also provide net carbon storage. The above results indicate the large net benefits that could be attained by reversing the impacts of historic drainage and degradation. Just halving the water-table depth in peatlands that are currently used for grassland (from 50 centimetres to 25 centimetres) and cropland (from 90 centimetres to 45 centimetres) globally would reduce the total global land-use related emissions of carbon dioxide equivalents by 11.5 per cent. Therefore, focussing on peatland management as a key climate change mitigation strategy seems essential.

In addition to changing our energy, land-use and industrial systems (to tackle all parts of Figure 4.6), we may also need to undertake geo-engineering to capture carbon and reverse global warming. There are hundreds of potential geo-engineering techniques being considered and tested, and there are already working technologies for capturing carbon from power stations and other industrial plants and pumping it into natural underground gas stores or disused oil wells. Sites for carbon capture and storage include locations in the North Sea, at In Salah in the Algerian central Sahara or the Lacq gas processing plant, south west France. Fawzy et al. (2020) conducted a review of the various strategies. Those most discussed in the scientific literature are provided in Table 4.1, which also briefly explains what each geo-engineering strategy involves. Note that there are negative-emissions-focussed strategies (e.g. grinding up rock and applying it to soils to enhance rock weathering rates – see Box 4.3) and also radiation balance strategies. The latter try to alter the Earth's radiation balance without adjusting greenhouse gases (e.g. by reflecting more of the Sun's energy back to space). Most of the radiation balance geo-engineering techniques are either at very early stages of idea development and testing or are associated with high risks or costs in terms of large-scale deployment. Many geo-engineering techniques also face important ethical challenges.

BOX 4.3 GEO-ENGINEERING USING BASALT ON CROPLAND

Experiments are showing strong benefits of delivering negative emissions using enhanced weathering. For example, Kelland et al. (2020) reported on an addition of 10 kilograms per square metre of relatively coarse-grained crushed basalt to a UK clay-loam agricultural soil. This one-off addition significantly increased the yield (by 21 per cent) of *Sorghum* without the use of phosphorus and potassium fertilisers. The crops were also stronger and showed potential benefits in terms of crop resistance to biotic and abiotic stress. Analysis of the soil properties 1–5 years after the addition suggests that the system had carbon dioxide sequestration rates of 2–4 tonnes per hectare. This represented an approximately four-fold increase in carbon capture compared to the control systems without basalt additions. Further field trials in a range of different environments are required, but this technique seems very promising in supporting global food security and carbon capture at the same time.

Table 4.1 The main radiative forcing geo-engineering techniques that are discussed in the scientific literature according to the review by Fawzy et al. (2020). Note that several of these have not been deployed.

Technique	*Description*	*In use*
Negative emissions technologies		
Bioenergy carbon capture and storage	Energy crops or waste from agriculture is combusted to produce energy and the emissions are captured and stored in geological reservoirs.	Yes
Biochar	Crops and agricultural residues are heated in the absence of oxygen to produce a carbon-rich char that is stable and is difficult to further break down. The char can be applied to soils.	Yes
Enhanced weathering	When silicate rocks weather through rainfall, carbon dioxide is taken up (see Chapter 2). The process is enhanced when the rocks are ground up and the ground up material can be added to croplands.	Yes
Direct air carbon capture and storage	Chemical bonding by sorbents (e.g. potassium hydroxide) used to extract carbon dioxide directly from the air. The carbon dioxide can be released from the sorbents by heating for release into geological reservoirs or to utilise it for other purposes such as chemical production.	Yes
Ocean fertilisation	Adding nutrients such as phosphorus, nitrates and iron to the upper ocean to enhance carbon dioxide uptake by promoting biological activity.	No
Ocean alkalinity enhancement	Carbon dioxide uptake can be increased in the oceans when the pH is increased. Ground up alkaline rocks could be added.	No
Soil carbon sequestration	Enhancing land management practices to encourage more carbon to be stored in soils, such as zero tillage and perennial cropping systems.	Yes
Afforestation and reforestation	Tree planting in new areas or where historic deforestation took place.	Yes

Technique	*Description*	*In use*
Wetland construction and restoration	Wetlands, such as peatlands, are major carbon stores and their protection and restoration promotes net carbon uptake. Storage, uptake and potential release depends on wetland type.	Yes
Radiative balance engineering		
Stratospheric aerosol injection	Reflecting aerosol particles are released into the stratosphere from specialist planes.	No
Marine sky brightening	Seeding clouds over oceans with sea-water particles or chemicals – the small droplets easily evaporate, leaving behind salt crystals that increase low-altitude cloud reflectivity.	No
Cirrus cloud thinning	High altitude cirrus clouds block long-wave radiation escaping back out to space and so thinning these clouds could allow more radiation to escape. Aerosol seeding has been proposed to speed up ice crystal formation and remove water vapour, thereby shortening the time for which the cloud is present.	No
Space-based mirrors	Reflectors transported into orbit around the Earth to block radiation from the Sun.	No
Surface-based brightening	Painting urban surfaces white and putting out reflective sheets across deserts.	No

CLIMATE CHANGE ADAPTATION

As well as mitigation, adaptation is needed to deal with the inevitable impacts of climate change that will occur before mitigation measures take effect. For example, adaptation measures include moving people away from floodplains that are more likely to be flooded in the future or designing better disaster management and emergency response procedures. In Europe, the health and economic impacts of heat waves are being dealt with by developing warning systems; altering dwellings, transport and energy infrastructure; reducing emissions to improve air quality; and having better wildfire management plans. Most governments are developing climate change adaptation policies, including around water management, coastal management and land planning. Infrastructure protection is another key aspect of climate change adaption management. In regions such as Central and South America, the use of protected areas, conservation agreements and community management of natural areas is occurring to support ecosystem adaptation. Communities in the Arctic are having to rapidly adapt to changing sea ice and permafrost patterns, which is altering food-sharing networks among communities and damaging infrastructure.

SUMMARY

- Many parts of the global carbon cycle have been changed by human action over the past few hundred years, with very rapid changes in the most recent decades.
- Climate changes over the past 2.4 million years have been dominated by cycles of cold and warm, during which major ice sheets expanded and contracted again.
- The driver of historic climate changes over the Quaternary has been orbital forcing, although this is not sufficient in itself to explain the magnitude of changes that have been experienced.
- Internal feedback processes are important for reinforcing or dampening climate change; of particular importance are the ocean circulation systems.
- The concentrations of greenhouse gases, such as carbon dioxide and methane, in the atmosphere are greater than at any time in at least the last 800 000 years, and their concentrations have risen sharply in recent decades.
- Evidence for the human influence on recent warming of the climate system is compelling, and since the 1950s many of the observed changes are unprecedented over millennia.
- The 2015 Paris Agreement saw more than 100 countries commit to achieve net zero emissions by the middle of the twenty-first century in a drive to keep global warming below 1.5°C since pre-Industrial times. Many businesses and individuals have committed to achieving net zero more immediately.

- The impacts of contemporary climate change will be far reaching, costly and hazardous.
- Mitigation measures need to be accelerated, including the use of low carbon energy sources and facilitation of enhanced carbon capture.
- Adaptation measures need further investment and development now to deal with the impacts of climate change that will happen as a result of emissions we are already committed to.

FURTHER READING

Cronin, T. (2010) *Paleoclimates: understanding climate change past and present*. New York: Columbia University Press.

A clear textbook outlining evidence and methods for evaluating change in the Earth's climate in deep time, the Quaternary and more recent periods.

Evans, C.D. et al. (2021) Overriding water table control on managed peatland greenhouse gas emissions. *Nature, 593*, 548–552.

A research paper that provides more information relevant to the finding described in Box 4.2.

Holden, J. (ed) (2017) *An introduction to physical geography and the environment* (4th edition). Harlow: Pearson Education.

Focus on three, well-illustrated, textbook chapters by specialists covering more detail on 'The Pleistocene (pp. 81–107), 'The Holocene' (pp. 108–136) and 'Contemporary Climate Change' (pp. 175–194).

Hubau, W. et al. (2020) Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature, 579*: 80–87.

The research paper that provides more information relevant to the discoveries outlined in Box 4.1.

IPCC (2014) The IPCC climate change reports are updated every few years and the latest versions can be found at http://www.ipcc.ch. In particular, the synthesis report is useful: IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the

Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. Geneva, Switzerland: IPCC.

5

WATER AND ICE

WATER

THE WATER CYCLE

Water evaporates from oceans, soil, living things (transpiration), rivers and lakes. At some point this water vapour condenses to water or ice in the atmosphere and returns back to the Earth's surface as precipitation. Some of this precipitation will land on oceans (~ 385 000 cubic kilometres per year) and some will fall on land (~ 100 000 cubic kilometres per year), where a fraction will infiltrate into the soils and rocks below the land surface where it flows more slowly to river channels or, sometimes, directly to the oceans. The water remaining on the surface and in the upper layers of the soil will partly evaporate back to the atmosphere and partly run into rivers and lakes. Water will eventually evaporate from inland flowing rivers along their course or at their lake exit point, while other rivers flow into the oceans (~ 39 000 cubic kilometres per year). There are four main stores of water. These are the world's oceans, polar ice, terrestrial waters and atmospheric water. The oceans hold 93 per cent of the water (~ 1 338 000 000 cubic kilometres); polar ice holds 2 per cent; the soils, lakes, rivers and groundwater hold 5 per cent; and the atmosphere holds a thousandth of 1 per cent of water resources. The water held in glaciers and polar ice, or deep within some rocks, may be stored for several thousand years. The water held by plants may just be stored for a few hours.

DOI: 10.4324/9781003124405-5

WATER MOVEMENT THROUGH THE LANDSCAPE

Precipitation can hit the land surface or be intercepted by vegetation. Some snow may undergo sublimation (evaporate directly from solid form) and return to the atmosphere, some may be stored in glaciers or ice sheets, and the rest may melt and flow into rivers and groundwater systems. Some intercepted rainwater on plants can evaporate, while some can flow down plant stems or drip from leaves in order to reach the land surface. The water that reaches the land surface can either infiltrate into the soil or pond up and flow over the surface as overland flow.

Infiltration is influenced by vegetation cover, soil texture, soil structure, the amount and connectivity of soil pore spaces and surface compaction. The **infiltration rate** of soils is often measured, which is the volume of water passing into the soil per unit area per unit time. The maximum rate of infiltration when there is a plentiful water supply is the **infiltration capacity**. The infiltration capacity of a soil generally decreases during rainfall. Therefore, if rainfall is at a constant rate, then the water arriving at the soil surface may at first all infiltrate into the soil, but then, as infiltration capacity decreases, more of the water will run over the land's surface. Water moving over the land surface tends to get into rivers more quickly than water that infiltrates into soil and moves within the soil. Compaction of the soil, surface crusts or a frozen surface can restrict infiltration, even if within the soil itself water percolation rates are quite fast. These restrictions to infiltration can lead to fast generation of overland flow and potentially large river flood peaks. Soils with lots of humus and a deep litter layer, such as those within tropical rainforests, tend to have large infiltration capacities.

If surface water supply is greater than the rate of infiltration then overland flow can begin once small surface depressions begin to overflow. This process is called **infiltration-excess overland flow**. This type of flow is uncommon in many temperate areas, except in urban locations, along roads and paths, or perhaps along compacted soils in arable fields created by tractor wheels or overgrazing by animals. Infiltration-excess overland flow is more common in semi-arid regions where soil surface crusts have developed and rainfall rates can be rapid. It is also more likely in areas where the ground surface is often frozen, such as northern Canada or parts of Siberia.

There is another important overland flow process known as **saturation-excess overland flow**. When all the pore spaces are full of water then the soil is said to be saturated and the water table (highest point below which the soil or rock is saturated) is at the surface. Water can therefore leave the soil and run out over the surface. This is common in shallow soils or at the bottom of hillslopes where water running through the soil will collect (Figure 5.1), keeping the pore spaces full of water for long periods and allowing water to return to the surface at this point. This means that saturation-excess overland flow can occur long after it has stopped raining. If it is raining then this additional rainwater will find it difficult to enter the soil if it is already saturated, so saturation-excess overland flow can be a mix of fresh rainwater and water that has been within the soil for some time.

The landscape area that produces saturation-excess overland flow varies through time. During wet seasons a larger area of soil will be

Figure 5.1 Saturation-excess overland flow accumulating along the bottom of hillslopes many hours after rainfall has stopped. Note here that because the fields are cropped, recent low crop cover has meant soil erosion has occurred and small rills have formed where overland flow has become concentrated.

saturated and able to generate saturation-excess overland flow than during dry seasons. If the catchment starts off relatively dry then during a rainfall event not much of the area will generate saturation-excess overland flow, but as rainfall continues more of the catchment will become saturated, especially in the valley bottoms, and therefore a larger area of the catchment will produce saturation-excess overland flow. This means that the source areas for saturation-excess overland flow are variable, whereas the source areas for infiltration-excess overland flow tend to be the same between similar-sized storm events.

Water moving through soils or rocks is called **throughflow** (sometimes also called interflow). Most rivers around the world are mainly fed by throughflow. This process can maintain river flow during dry periods (the **baseflow** of the river). There are different ways water can move through soil or rock (Figure 5.2), which affect how quickly water can get to a river channel and therefore the typical nature of the response of a river to rainfall. **Matrix flow** occurs when water moves through the small pore spaces within a rock or soil, whereas in **macropore flow** water moves through larger cracks, fissures or continuous channels within the rock or soil, thereby bypassing contact with most of the soil mass itself. Definitions vary but macropores are often considered to be pores larger than 1 millimetre in diameter, although to be hydrologically effective they need to form connected conduits for flow (Figure 5.3). It is possible to estimate the potential velocity of water flow through the matrix of a saturated soil or rock (known as the **hydraulic**

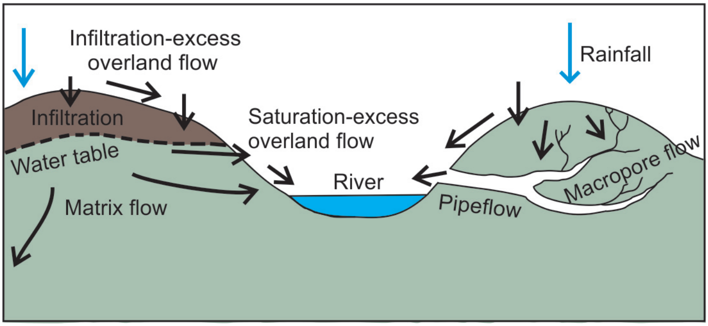

Figure 5.2 Key hillslope flowpaths.

Figure 5.3 A collapse feature showing the subsurface route of a natural soil pipe with flowing water.

conductivity, measured in metres per second) using techniques such as dye tracing, pumping water out of a well and timing the length of time it takes to rise back up the well again, or taking small samples back to the laboratory and running water through them. However, even if a soil or rock is very permeable, and therefore has a high hydraulic conductivity, the actual rate of flow of water through the substance will also be determined by the **hydraulic gradient**, which is the difference in hydraulic pressure between two points divided by the distance between those points. In other words, if the soil is saturated throughout a steep hillslope there will be a larger hydraulic gradient across the slope than if the soil is saturated all along a gentle gradient hillslope.

When water moves through the soil or rock matrix there are good opportunities for cation exchange (see Chapter 2 on soil chemical properties) and changes to the chemistry of the water. However, some soils have lots of macropores within them, many of which actively transport water. At some sites measurements have suggested more than half of the water moving through soils could move through macropores. If an arable field has many macropores, surface fertiliser applications may get washed through the macropore channels and may not enter the main part of the soil. In this situation, the fertiliser will not be available for crop use and may enter the stream or deeper groundwater system, leading to water quality problems. Macropores can be formed by soil animals, plant roots and cracking during dry conditions, or from small landslips. Some macropores can be tens of centimetres in diameter, others can be metres in diameter, which are often known as natural soil pipes (Figure 5.3). Soil piping is more common in erodible soils, such as on the Loess Plateau of China, in semi-arid southeast Spain or Arizona or in karst limestone cave systems, but has also been well documented in some types of wet peatland systems. Turbulent flow within these large macropores in soils can rapidly enlarge them further, until eventually they collapse to form gullies.

GROUNDWATER

The term groundwater can be defined in many different ways. Some view it solely as water within rock. Another working definition is that groundwater is water held below the water table, both in soils and rock. In any case, water held within the ground is of worldwide importance. In many catchments water is supplied to the stream from groundwater in the bedrock. The groundwater has percolated down through the overlying soil and entered the bedrock. Rock has small pores, fractures and fissures. Groundwater stores around 30 per cent of the Earth's freshwater. However, if this groundwater is to be available to supply river flow, the rock or soil needs to be permeable, enabling water to flow through it. Layers of rock porous enough to store water and permeable enough to allow water to flow through them in economic quantities are called **aquifers**.

More than a fifth of the world's population rely solely on groundwater to satisfy their needs. Some countries, such as Hungary and Denmark, almost entirely rely on groundwater for their drinking water supplies. France obtains 56 per cent of its drinking water from groundwater, while the UK obtains around 30 per cent from this source. The USA obtains around half of its water for all uses from groundwater, with a quarter of drinking water coming from aquifer resources. However, a greater proportion of groundwater is served to communities with smaller populations, which makes quality control more difficult (e.g. from a small well supplying 100 homes). People have obtained groundwater for thousands of years through digging wells or collecting water from springs, which are zones of groundwater emergence. However, if the amount of water being withdrawn from the ground by humans is not replaced at the same rate by inputs to the aquifer, then groundwater levels will fall (see Box 5.1). Indeed, a rise and fall in the water table naturally happens seasonally in many areas depending on rainfall and evapotranspiration rates. However, in many places groundwater has been greatly depleted, particularly where rates of replenishment are slow and water that is being abstracted may be hundreds or even thousands of years old. Around a tenth of global groundwater consumption is therefore thought to be unsustainable. The costs of pumping out deeper water are greater than obtaining water near the surface. In some places, subsidence has resulted due to groundwater depletion, such as in Mexico City or Tuscon, Arizona. In the latter, some locations experienced over 8 metres of subsidence, until groundwater use was reduced by the construction of a long canal to divert surface river water 400 kilometres across the desert from the Colorado River into Tuscon. Because of the links between groundwater and baseflow, rivers can dry up if groundwater is over-abstracted. In coastal areas, where the aquifer is directly connected to the sea, groundwater abstraction can lead to the intrusion of salty water to replace the lost groundwater under land. This is a problem, for example, along lots of Australian coastlines where there are major cities, and on the shores of the Persian Gulf.

Groundwater management is therefore a key priority across much of the world, and now even more so given the need to adapt to climate change. Such management requires good information on rates of inputs and outputs of water and trying to understand the geological 'boundaries' of a groundwater supply area. Armed with this information it should be possible to evaluate possible management strategies. These strategies may also need to take into account the local needs of the environment (e.g. outflow of water that supplies river baseflow and maintains aquatic **biodiversity**).

Artificial recharge of aquifers is being attempted in several places, such as Long Island, New York, and the Dan Region, Israel. These practices encourage infiltration and direct injection of water during wet periods via pumping wells. They may also involve creating a strong hydraulic gradient (e.g. by pumping water out of the system to other storage areas in order to encourage water ingress into the groundwater from adjacent lakes). Water efficiency measures are often also required to reduce water demand, including for irrigation purposes, though this can be challenging given growing populations and food security concerns.

BOX 5.1 SATELLITE-DERIVED GROUNDWATER DEPLETION ESTIMATES FOR THE MIDDLE EAST

In many parts of the Middle East there is limited surface water availability to supply human populations and so groundwater abstraction is dominant. Such abstraction has been underway for millennia, but with burgeoning populations and enhanced economic development, demand for water has been increasing. In fact, of the world's ten countries with the least renewable water resources per inhabitants, seven of them are Middle Eastern countries: Kuwait, United Arab Emirates, Qatar, Saudi Arabia, Yemen, Bahrain and Jordan. Given the large area it is quite difficult to assess groundwater depletion across the region based on borehole data and local observations. However, satellites have been used to make an assessment of the changes to regional groundwater. The Gravity Recovery and Climate Experiment (GRACE) satellite mission measures the gravitational forces on Earth. As the mass of water in the groundwater system declines, the gravitational force weakens. One study (Voss et al., 2013) using GRACE data for part of the Middle East has shown a very rapid decrease in total water storage of 91 cubic kilometres of water over just 7 years. The loss of such water in a water scarce region means that additional costs and energy use are required to pump deeper water from the ground and also to obtain water from other sources, such as by removing salt from seawater, which can be very energy intensive.

RIVER FLOW

Water flowing through and across the land surface is called runoff (note that runoff does not just mean overland flow). Runoff often flows into rivers or lakes. The area of land that could potentially drain into a river or lake is known as the catchment area or watershed. Rivers can be fed by subsurface throughflow or by overland flow, and the relative proportions of the different types of flow can determine how quickly the river flow changes during rainfall events or seasonally. The proportion of rainfall falling on a catchment which then reaches a river channel may vary from close to 100 per cent to 0. River flow is crucial for aquatic life, water availability for reservoirs, abstraction for human use and riverine flooding.

River flows can change during individual rainfall or melt events, or remain fairly stable, depending on the nature of the environment. There tends to be a lag between the precipitation occurring and the peak discharge of a river. This lag time is affected by the runoff processes discussed above. Where infiltration-excess overland flow dominates the runoff response then the hydrograph is likely to have a short lag time and high peak flow. The hydrograph (a graph of river discharge through time) will therefore be quite steep (perhaps even spiky) in shape, rising quickly from low flow to the peak flow. However, if the catchment area is very large (tens of thousands of square kilometres) then it may still take several days before peak river flow occurs in the downstream reaches. Urbanisation increases flood risk as it reduces the infiltration capacity of the surface through construction, leading to rapid runoff to the river, which results in steeply rising hydrographs with sharp peaks. If throughflow moving via the pore spaces in soils and rocks (matrix flow) dominates runoff response then the river may rise and fall very slowly in response to precipitation and the peak may be small. However, since throughflow contributes to saturation-excess overland flow then throughflow can still lead to rapid and large flood peaks. In some soils such as peat, only a small amount of infiltration may be needed to cause the water table to rise to the surface. In other soils there may even be two river discharge peaks caused by one rainfall event. This might occur where the first peak is overland flow dominated and the second peak, a little while later, which may be much longer and larger, is caused by subsurface throughflow accumulating at the bottom of hillslopes and valley bottoms before entering the stream channel. Throughflow may also contribute directly to storm hydrographs by a mechanism called piston or displacement flow. This is where soil water at the bottom of a slope is rapidly pushed out of the soil by pressures from upslope caused by new, fresh infiltrating water entering at the top of a slope.

Figure 5.4 shows flows in the form of hydrographs over one year for two nearby rivers where the climate is the same. Despite being in the same area the flows are very different between the rivers. The flows in the River Nol appear to be dominated by baseflow and there are no individual storm peaks, unlike for the River Creef which appears to be more dominated by overland flow or rapid throughflow (e.g. via macropores and soil pipes). The River Nol catchment overlies permeable bedrock, is gently sloping, and has soils that enable good infiltration and little chance of saturation to the surface. Therefore, infiltration-excess or saturation-excess overland flows are a rare occurrence here. For the River Creef, however, the soils are thin and

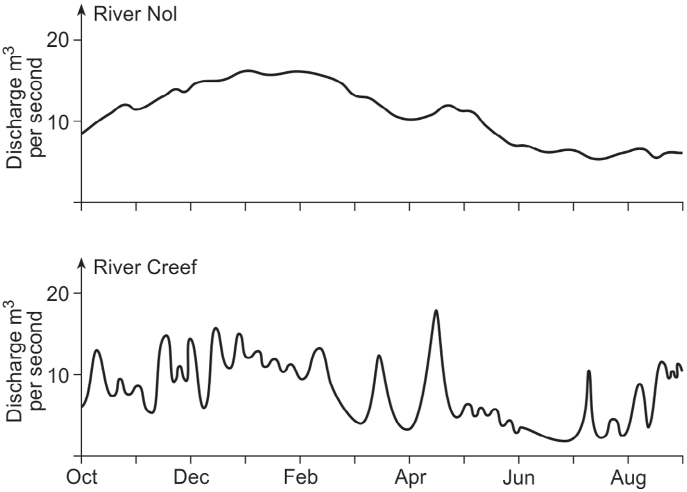

Figure 5.4 The changes in river discharge over one year for two nearby rivers with similar rainfall but different catchment characteristics.

sit above impermeable bedrock, and so there is frequent saturation of the soils and generation of saturation-excess overland flow.

Seasonal variability in river flow is known as the **river regime**, for which there are four major global types. Where snow and ice melt dominate there can be a major peak of river flow during the late spring, in the case of snow melt, or early summer, with annual glacier melt. River discharge (the volume of water passing a point in the river per unit time) can be extremely low during the winter months even though precipitation may be continuing, as this precipitation is stored on the glacier in winter. There can also be a strong daily change in river discharge due to daily melt cycles of the snow and ice. Night-time discharge in cold regions tends to be much lower than that of the mid-afternoon.

Arid zones, especially in subtropical drylands, tend to experience very occasional but intense rainfall events. Intense rainfall along with little vegetation cover produces infiltration-excess overland flow, rapid runoff and high flood peaks. However, many dryland soils are coarse and sandy with high infiltration capacities, resulting in little chance of overland flow. Therefore, there is a wide variation of response even if rainfall intensities are very high. In most drylands, river flow will stop within a few days of the rainstorm and water often seeps into the river beds or is evaporated.

In temperate, oceanic areas, precipitation occurs periodically all year, perhaps with seasonal maximums (see Chapter 3). The river flow regime in these areas can change in response to either seasonal changes in groundwater storage and release, or higher evaporation and transpiration rates during the summer months.

Rivers in equatorial areas tend to have a fairly regular regime, while tropical river systems outside of the equatorial areas receive high precipitation during the summer but experience a marked dry season during the winter. Evaporation and transpiration is high at all times, so that the streamflow mirrors the seasonal pattern of rainfall.

Land and river management can dramatically alter the flows in river systems. Building dams and changing or diverting stream courses alters river regimes. In fact, many major rivers are now unnaturally dry for at least part of the year at their outlets to the sea or inland lakes because of dams and water abstraction. These include the Colorado (Mexico/USA, Figure 5.5), the Indus (South Asia) and the Murray-Darling rivers (Australia). Lack of flow at the

Figure 5.5 Landsat 8 image of part of the dry Colorado River channel and floodplain near San Luis Rio Colorado (Mexico)/San Luis Arizona (USA). The Canal Alimentador Central running roughly northeast-southwest can also be seen, which is one of the many abstraction routes for water away from the river. In the past, the Colorado would have flowed with high discharge through this area towards the Sea of California. However, so much water is withdrawn from the river that the connection to the sea is now rare.

(Source: NASA's Earth Observatory, https://earthobservatory.nasa.gov.)

outlet causes inland lakes to shrink or dry up where rivers flow into continental interiors (e.g. Lake Chad is now only a tenth of its former size), or for ocean-flowing outlets, the lack of discharge causes erosion of deltas because of a lack of sediment supply from upstream. Deforestation or intense grazing may result in a large reduction of the infiltration capacity of a soil and decrease in transpiration rates. Covering more of the landscape with concrete and tarmac, which is impermeable, and then channelling flow into drains that feed streams may lead to increased flood risk downstream. These processes lead to more water flowing to rivers with shorter lag times and therefore potentially higher flood peaks. A recent examination of data from urban streamflow gauges in the USA for small catchments of 5–162 square kilometres, where there was at least 20 years of data and no dams, has shown that high flows increased with urbanisation, whereas low flows increased in some places and decreased in others (see Bhaskar et al., 2020). This latter effect may relate to how, as an urban area grows, there can be water abstraction upstream, reducing low flows in rivers, but then downstream wastewater effluent can be released back into the stream, so that streamflow regime can vary depending on the design of such schemes.

Flooding is a natural phenomenon and should be expected. Each year flooding causes hundreds of deaths around the world and significant impacts on infrastructure. Floods commonly occur when rivers overtop their banks. However, flooding can also occur through a very high tide at the coast, often made worse by a storm event (**storm surge**) and high river flows. Flooding may also happen when very heavy rainfall cannot escape from an area (e.g. an urban area where the drainage system cannot cope with the excess surface water) or when there is a lot of saturation-excess overland flow. Flooding can also bring benefits to farmland through supplying nutrients. There is generally always a flood occurring somewhere in the world (see https://floodobservatory.colorado.edu/ for maps of current floods).

Flood risk is often interpreted by examining the historic frequency of flooding. If a water level greater than 10 metres in height occurred five times in the past 10 years we would say that the return frequency of the 10-metre flood is, on average, once every 2 years. However, this does not mean that a 10-metre-high flood will occur once in every second year, as the return frequency is based on a long-term value. It may be, for example, that in one year a 10-metre-sized flood occurs three times. In addition, there is concern that looking at past records might not be the best guide to predicting future flooding, particularly if land management change has taken place in the catchment or if climate change is likely to change the precipitation patterns or vegetation cover. Machine learning has recently been applied to flood data to support real-time flood warning systems. These techniques mean that as landscapes change, the machine can keep learning about the flood response and modify its predictions for residents and businesses (see Box 5.2).

Often our response to flooding has involved building larger hard flood defences around our towns and cities. However, the flood water has to go somewhere and sending the water more quickly through one part of the river system by building large embankments or straightening and deepening river channels simply reduces the lag time downstream and increases the overall flood peak there. There has therefore been a recent drive to take a more 'nature-based' approach to flood management. In urban zones this may mean adopting more green (planted) roofs or ensuring roads and pathways are permeable, in combination with creating ponds and lakes to temporarily store rainwater. Such approaches have been adopted in many regions, often as part of sustainable urban drainage systems (SUDS), which may also reduce pollution entering watercourses. However, much more ambition is required, such as the Sponge City concept in China where megacities (e.g. Ningbo) have been chosen to adopt radically intense water-sensitive infrastructure. However, it is also important to think beyond urban areas and to take a whole catchment approach. 'Natural flood management' is becoming a popular concept in some regions, such as Europe, where techniques such as tree planting, in-stream woody dams, soil management, farm ponds and revegetation are adopted across upstream catchment areas to try to slow water flow and reduce flow peaks to downstream areas. Data is still being collected on the relative success of such schemes as most evidence is based on datasets collected over fairly small-scale catchment areas.

BOX 5.2 USING BIG DATA AND ARTIFICIAL INTELLIGENCE TO TACKLE FLOODING

The city of Hull in England is a low-lying coastal town on the banks of the Humber estuary. The Humber river basin drains around one third of the area of England. Hull has been subject to combined flooding sources, including stormwater overflow from overwhelmed urban drainage systems, coastal flooding, pluvial flooding from heavy rainfall onto impermeable urban surfaces and riverine fluvial flooding.

Recently, working with the 'Living With Water Partnership' in the city, high resolution rainfall data and data on water levels in rivers and drain systems provided by several organisations has been put together for the first time into one combined data system. As part of a UK Natural Environment Research Council funded project, called the Integrated Catchment Solutions Programme, machine learning techniques were then applied to the resulting big dataset. Machine learning is a form of artificial intelligence whereby computers learn from patterns in data. Within vast volumes of data many of these patterns are not apparent to human interpretation. The computer algorithms found several key patterns related to connections between rainfall in certain locations and water levels a few hours later in different parts of the drain and channel network, and could also pick out tidal effects. As a result, it was possible for the computer to train itself to predict how different parts of the complex water network would respond to different rainfall patterns across the area. It is also possible for the machine to keep learning so that as the system changes over time (e.g. new housing developments, changes to drain networks) the computer can modify its predictions. Data are collected in real time via the measurement devices using mobile phone and wired networks deployed by the partnership of organisations. These data are run through the artificial intelligence system, which can predict the water level response in different locations, resulting in a real-time warning system. Using a special form of data map (not the same as a spatial map) it is possible to show the relationships between datasets in different parts of the water system. The system is triggered by rainfall patterns and allows residents and businesses, in the few minutes or hours before the effects are felt, to take actions in their neighbourhood to reduce the impacts (e.g. clearing debris from drains, moving cars to higher ground, closing flood gates, evacuating an area). Such an approach allows major advances in local flood management, and by using the data maps it also overcomes one of the key limitations of some machine learning tools as it also aids our understanding of why the output looks as it does from the large volumes of input data.

For more information about this case study and wider initiatives on flood management in the Yorkshire region of England, including natural flood management, communicating flood risk and flood forecasting, please see: www.icasp.org.uk.

RIVER CHANNEL CHANGE

Rivers are dynamic in that they move position around the landscape through time, change their shape and move sediment, water and dissolved materials. They are an important agent in removing weathered material from the land surface and redistributing it across the landscape or into oceans. This process balances out the mountain building caused by plate tectonics, so that over long time periods plate tectonics might build mountains but weathering, erosion and removal by river systems or ice masses (see later in this chapter) smooth out the landscape again. Uplift and erosion of landscapes has traditionally been thought of as cyclical, although the outcomes in each environment may be very different depending on rates and types of processes operating there.

Most rivers have a long profile (slope of a river from its source to mouth), which is concave with progressively lower gradients downstream. The long profile varies with geology, tectonics and variability in runoff. Other profile shapes also occur when there is interruption by lakes or very resistant rocks (which often result in waterfalls), or through large changes in sea level. If sea levels fall then the whole river may start to erode its bed downward in response. If sea levels rise the river may deposit more sediment lower down its course.

Sediment movement in rivers is linked with water flow. If a particle of sediment is to be picked up from the river bed or bank by flowing water a critical threshold has to be passed, above which the water velocity or **shear stress** is sufficiently large to overcome frictional forces that resist erosion. The transport of materials close to the bed is known as bedload transport. Particles move by rolling, sliding or saltating (hopping) along or close to the bed. If the flow velocity does not change, a particle will only come to rest if it becomes lodged against an obstruction or falls into an area sheltered from the main force of the water by a larger particle. With further increases in the strength of flow, the smaller particles may be carried upwards into the main body of water and transported in suspension. Deposition and cessation of movement for an individual particle occurs when velocity falls below critical conditions. This means that finer particles are preferentially moved downstream. For suspended sediment within the water body (as opposed to that transported close to the river bed as bedload transport), transport is determined not only by water flow but also by the rate at which it is supplied to the river (e.g. from wash processes – see Chapter 2).

Where erosion exceeds deposition within a particular section of a river, there will be lowering of the river bed or widening of the banks. If erosion and deposition occur at the same rate then the river channel will stay at roughly the same level. Eroding channels may undermine structures such as bridges, while depositing channels may submerge structures such as roads. Stable channels, especially those whose beds are lined with bedrock, are less likely to be a problem to engineering structures, but fluctuations in river channel dimensions and locations caused by flooding or sediment pulses moving down the river can be problematic.

River forms are often described by features of the cross-section of the channel. Natural channel cross-sections adjust to accommodate discharge and sediment loads. Channels are expected to be wider and deeper if the discharge is greater. Moving downstream, river channel cross-sections tend to get larger, although there is not a perfect relationship between discharge and channel cross-sectional area because larger channels are more efficient at carrying water (less friction around the channel edges per volume of water). Furthermore, the sediment that makes up the river channel is important. Channels with a high percentage of silt or clay in their banks (which are often more characteristic of lowland sections of river), and rivers transporting much of the sediment load in suspension, tend to be narrower and deeper than sand and gravel-bed rivers. Vegetation can also be important in controlling cross-section shapes by influencing bank resistance through root systems that bind the sediment. Removing bankside vegetation can lead to rapid bank erosion.

Within channel cross-sections there can be large variations in the velocity of water flow. Close to the bed or banks of the river the velocities tend to be slower. On bends of rivers there can be forces exerted that increase the pressure on the outer bank as the water flows by, with less pressure on the inner bank of the bend. Bank erosion is more likely on the outer bank as faster water has more chance of picking up sediment. On the inner bank, sediment is more likely to be deposited. This causes further development of the meander (Figure 5.6) and means that the river is continuously

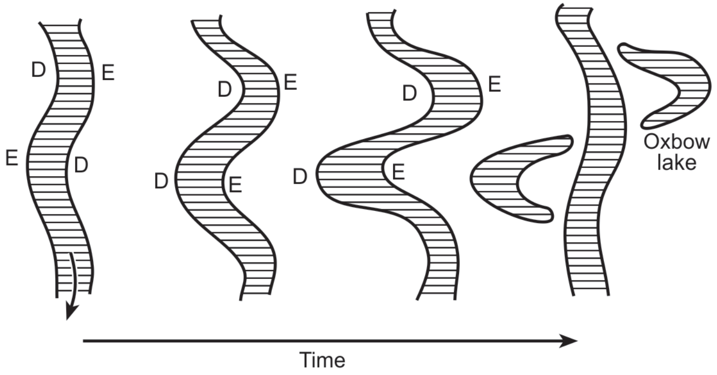

Figure 5.6 The enlargement of river meanders and formation of oxbow lakes. E is a zone of net erosion and D of net sediment deposition.

eroding and depositing sediment as a natural process, and so the exact position of the river bank will change through time. Sediment size tends to be greatest towards the outer bank of meandering channels, as water velocity is greatest close to the outer bank, and slowly decreases towards the inner bank. When meander bends become too exaggerated the river will cut through the meander to a direct course downstream for that short section, leaving behind a crescent-shaped section of channel that is cut off from the main river, known as an oxbow lake. This change in course for the river often happens during a high-flow event when erosion and deposition are most active. Sometimes, during such events, the river can abandon long stretches of its previous channel completely and take a new course on a different part of the floodplain. A sudden change in channel location is known as channel **avulsion**.

Looking at a river from above, the pattern, or planform, can be seen. River patterns occur mainly as braided, meandering and, very occasionally, as straight channels (Figure 5.7), although the latter is most likely to be human-made. A single river may have each of these patterns in different sections. Braided channels consist of lots of individual smaller channels, which separate and come together with islands between, known as bars. The channels can rapidly move locations with bars eroding on one side and growing on the other. Braided rivers are common where there is a lot of mobile

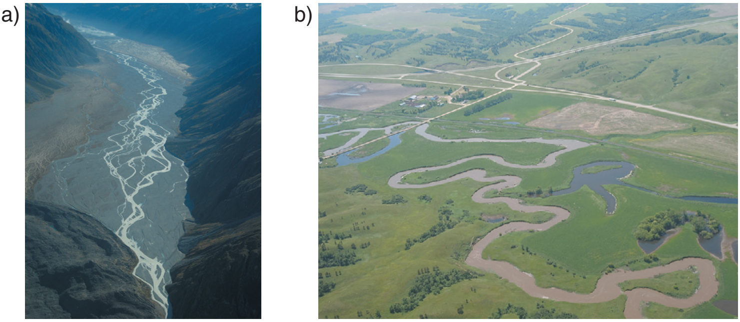

Figure 5.7 Classic river planforms: a) braided Murchison River, New Zealand; b) meandering Des Lacs River, North Dakota, USA.
(Source: a) Avenue, CC BY 3.0: https://creativecommons.org/licenses/by-sa/3.0; b) US Geological Survey, photo by Joel M. Galloway, https://www.usgs.gov/media/images/des-lacs-river.)

sediment, such as downstream of a glacier, and are essentially formed of many different meandering parts. Meandering channels have a wavy planform. Often, wavyness is measured by sinuosity to determine the nature of the measuring channel, where sinuosity is the ratio of the length of river between two points compared to the length of the straight-line valley between these two points. Straight channels are defined as having a sinuosity of less than 1.5. A meandering channel refers to a single channel with a number of bends that result in a channel sinuosity in excess of 1.5. Straight rivers are more controlled by human action, and natural straight rivers are often unstable and become meandering.

Channel slope and discharge are important controlling variables on the platform of river channels. For a given discharge there is often a critical slope above which channels will meander, and then a further threshold above which they will braid. These thresholds decrease with increasing discharge. Thus, braided sections are usually found on large rivers or on small rivers with steep slopes. Wandering and braided rivers will occur where there is coarse sediment, erodible banks and where the main sediment transport mechanism is bedload transport rather than suspended transport.

The characteristics of river beds tend to change downstream. Often there are bedrock channels in the upper section of a river network or large boulders and cobbles. There is a sharp decline in the size of bed material downstream. This pattern occurs because smaller particles are transported downstream more easily and because abrasion of larger sediments within the river by colliding and grinding causes material to get smaller and more rounded downstream. However, these patterns are not seen everywhere and may be disturbed by local sediment inputs to the river network.

Within rivers there are several erosional and depositional landforms. In bedrock channels, potholes can be found, formed by the mechanical wearing and grinding of small particles enlarging an existing small depression or weakness in the rock; pressure changes due to bubble collapse in turbulent flow; and chemical weathering. Within rivers that have gravels on their bed, the most common landforms on the bed are sequences of pools and riffles. Pools are deep sections with relatively slow flowing water and fine bed material. Riffles are formed by accumulation of coarse sediment with shallow, fast flowing water. The spacing of pools and riffles is often five to seven times the channel width but the spacing does vary. The bed of sandy sections of rivers can have small ripples in the sand, which are less than 4 centimetres in height, and then also larger dune features. The size and shape of these dune features change with discharge following rainfall events. The dunes and ripples tend to migrate downstream as sand is carried up the upstream facing side of the ridge of the feature and then falls down the downstream facing side. At very high flow velocities a flat river bed can be formed or dunes can migrate upstream, since erosion from the downstream side of the dune allows suspension of material in the water, which occurs faster than it can be replenished from upstream.

River channels change their slope, cross section, planform and bed forms in response to environmental change. However, humans have modified river channels over the past few thousand years, with changes during the last two centuries having been the greatest. Activities such as dam construction, urbanisation, mining, land drainage and deforestation can all impact river channels. Faster, more peaky flow from urbanisation or deforestation can accelerate erosion. In the USA, channel volumes have been found at up to six times the size of those of similar rivers with more natural flows. Dams are having a major impact on river flows, channels and sediment dynamics. For example, the Nile now only transports 8 per cent of its natural load of silt below the Aswan Dam, thereby reducing the fertility of the downstream floodplains and accelerating river bed and coastal erosion as the lack of sediment entering the sea no longer replenishes the sediment being eroded by wave action (see coasts section below).

The above discussion has shown that river channels are not static features in the landscape. Despite this, rivers often act as boundaries for land ownership, which can lead to disputes because the river course changes through time and so the boundary between property, counties, states and countries can change too. Understanding the processes and dynamism involved in river channels is essential to good management, otherwise great expense might be incurred when engineering structures fail due to river channel change. There are many examples over the last century of engineering failures around rivers. A notable example of this is the Mississippi River, which regained much of its sinuosity following engineered channel straightening in the early twentieth century. Indeed, many rivers that have been subject to engineering features, such as straightening, became less diverse environments as the biological variety within the river was removed, so that ecological functions suffered. The biological diversity was supported by the diversity of channel features (e.g. pools, riffles, shade, slow moving and fast moving sections) and by the diversity of flows during the year. More uniform channels and uniform flows tend to be less biodiverse. Aside from the aquatic ecology, terrestrial ecology often also suffers because the two interact (e.g. birds feeding on aquatic species). Many rivers are now being rehabilitated to try to restore features that encourage greater wildlife (such as meander bends with pools and riffles, natural sediment rather than concrete) and also to work with the erosion and deposition processes within rivers rather than against them. For examples of river restoration projects that attempt to reintroduce natural platforms and channel features please see www.therrc.co.uk. One of the key challenges around this is that rivers naturally migrate their course over time, but humans like to create fixed infrastructure around water courses. Thus, restoration projects have to balance the recreation of natural processes alongside socio-economic demands for resilient infrastructure.

WATER QUALITY AND POLLUTION

Water quality is a measure of the various chemicals that are found within the water. Water quality can vary naturally, related to the local geology, soil types and climate, and it can be influenced by human action. Water pollution occurs when the water environment is changed so that the species using the water can no longer tolerate its chemistry and either die or quickly leave that section of the water body. Often, water quality and pollution are measured from a human perspective in terms of how safe the water is to drink or how easy or costly it is to treat the water before it can be consumed.

Water that has a good taste and is not dangerous to health still contains dissolved substances. The highest quality waters with the best taste tend to come from reservoirs and lakes that have collected the majority of runoff from undisturbed landscapes with mainly near-surface or overland flow. However, many groundwaters can taste good too. These tend to be where rocks only weather slowly. If groundwater has picked up lots of dissolved substances from soils and rocks it is more likely to have an unpleasant taste.

The runoff routes for water across and through the landscape are important for controlling the concentrations of different chemicals within rivers, lakes and groundwater. Precipitation inputs normally have low concentrations of dissolved chemicals. Therefore, where a site is dominated by rapid water movement to rivers (macropore flow or infiltration-excess overland flow) there is little time for chemical interactions with soil or rock, and so river water chemistry will be similar to rainwater chemistry (e.g. in blanket peatlands). However, if there is significant overland flow that brings lots of sediment with it, perhaps with other added fertilisers or industrial chemicals, then that can cause pollution problems. Throughflow in upper soil layers and water coming back out of the soil as part of saturation-excess overland flow will tend to have quite different solute concentrations to that of precipitation since it will have had more time to interact with the soil and for weathering reactions to take place. Solute concentrations in groundwater are often greater than elsewhere due to the longer contact time of water with soil and rock. The groundwater composition is affected by the geochemistry of the surrounding geology since different minerals weather at different rates and produce different solutes.

Water quality varies with time because the water flow processes across the landscape also vary with time. As groundwater and throughflow deep within the soil are the major source of base cations (see Chapter 2) to river water from mineral weathering, base cation concentrations in river water tend to be greater during dry conditions when groundwater is the main source of river flow. In wetter conditions more flow is generated from macropores and over the land surface, containing smaller base cation concentrations. The availability of solute supply can vary with temperature during the year or with plant growth (e.g. nitrate concentrations can be lowest in spring and summer due to plant uptake). The concentrations of chemicals within precipitation can also vary with time. In coastal regions, precipitation is often enriched by sea salt during stormy weather, increasing chloride concentrations in river waters.

Solute concentrations and **solute fluxes** (total mass of a solute moved) in rivers and lakes vary locally and globally. This variation is due to local and regional differences in climate, geology, topography, land management, soils and vegetation. Globally, patterns of solute concentrations are dominated by climate and geology. On a regional level, land management, soil types and vegetation are more important. Land use affects solute concentrations by altering runoff pathways and the amount and availability of dissolved chemical sources. It has been estimated that human action has increased the total amount of solutes transported by rivers across the planet by 12 per cent. Box 5.3 provides details of some recent progress linking satellite observations to water quality sampling to determine patterns of change.

BOX 5.3 CHANGES IN WATER QUALITY MEASURED FROM SPACE

The AquaSat dataset combines water quality sample data from rivers, streams and lakes in the USA with more than 30 years of remote sensing images taken by Landsat satellites. These images can detect changes in colour of water, which might be driven by sediment, dissolved organic carbon and chlorophyll *a*, which is a measure of algae that turns the water a greenish colour. Ross et al. (2019) correlated colour properties of water bodies from Landsat images between 1984 and 2019 with ground-based water sample results collected by US agencies over time, yielding 600 000 matching records between the two. This dataset now provides a calibration for remote sensing images so that satellite imagery can be used to estimate long-term change in water quality for water bodies that have not been monitored by sample collection in the past. Extending such work globally will be an important future exercise and may allow better spatial understanding of water quality changes in areas that have not been well sampled in the past.

Agriculture has had a large impact on water quality, particularly through increased erosion and the leaching of nutrients, pesticides and by-products of veterinary medicines into water courses. There may be individual point sources of chemicals from leakage of pesticides, slurry and wastes from storage facilities, or from more diffuse pollution across the landscape. Manure and slurry are often spread on farmland, but when the plants cannot take up all of the nutrients provided (i.e. too much has been applied) or if the manure is applied just before heavy rain then leaching of soluble nitrogen and other chemicals occurs. Drainage and ploughing can also increase leaching rates and erosion. Thus, there has been much effort to support agri-environmental measures that reduce pollution but also cut costs for farmers by ensuring soil and applied fertilisers stay on the land.

There have been many instances where groundwater contamination has been related to damage to human health (e.g. too much fluoride causing debilitating bone diseases or too much nitrate causing lack of oxygen in the blood; arsenic poisoning, which can be an acute problem in some regions, including in parts of Bangladesh and West Bengal, India). Often it is long-term exposure that is the problem rather than a one-off drink. This is because groundwaters, unlike surface waters, are often fairly clean from sources of infectious disease, but it depends on how good a filter the soil and rock are within the aquifer. Human health problems have led to legal standards for the concentrations of many solutes in drinking water within most countries. Contamination can take place quite some distance from the abstraction point if the aquifers are well connected. Surface landfills, chemical spills, food processing (e.g. Box 5.4), mining operations, leaking underground tanks at fuel stations and many other surface management strategies lead to contamination of groundwater from point sources. Diffuse sources, such as fertilisers and pesticides from farming, may also be important contaminants in some areas. In many places there is only a slow percolation of contaminants into some aquifers. For some deep aquifers, the water may take 50 years or more to percolate down. In some countries the peak use of fertilisers that were applied at rates that were too great for plant uptake occurred in the 1970s or 1980s. Here, despite good environmental protection measures since then, and careful controls over the use of fertilisers in groundwater source zones, the levels of solutes such as nitrate may be continuing to increase in groundwater. There is, therefore, a large time lag between surface management and groundwater quality changes in some locations, meaning the groundwater problem only emerges several decades after good management has been implemented. One solution to this is to mix the contaminated groundwater with water that has relatively low concentrations of solutes to enable all of the water resource to be used. However, this would require multiple water sources for a supply area, which is not always possible. Instead, very costly removal techniques have to be applied to strip out the contaminants from the water.

BOX 5.4 GROUNDWATER POLLUTION FROM TAPIOCA IN SOUTHEAST VIETNAM

Tay Ninh is a province in southeast Vietnam of around 4 000 square kilometres and a population of around one million. Tay Ninh has cultivation specialised in crops including sugar cane, peanuts and cassava. Factories associated with processing these crops are near the farms but are not close to surface water networks, and so use large amounts of groundwater drawn from deep aquifers. However, nearby households abstract groundwater from shallow aquifers using dug wells (~ 5–6 metres deep) and boreholes (20–25 metres deep). Most of the factories employ pond systems to treat high organic content wastewater from food processing. However, these pond systems were often not made with fully impermeable liners and so organic-rich wastewater from the food processing plants has percolated into the shallow aquifer used by local communities. Many of the groundwater wells in the region have previously been found to be highly contaminated. High levels of chemical oxygen demand, a key water quality indicator, was associated with cassava root processing to make tapioca, which releases high starch content and also cyanide. The Vietnam government has now stipulated that the factories must use impermeable pond liners, and new treatment plants have also been tested and installed to treat the wastewater. However, while the factories themselves have been allowed to insert very deep boreholes to extract cleaner water from deeper aquifers for their processing operations, there is still a legacy for shallow groundwater quality and it may be many decades before the groundwater zone previously used by residents provides water that is suitable for human consumption.

Urbanisation and industrialisation are associated with increased river concentrations of metals, nutrients, organic matter and even road salt in the winter in cold climates where roads are treated. Vehicles are responsible for a lot of urban metal pollution and microplastics as they corrode and wear, depositing materials on the urban surface. Urban drainage with rapid removal of surface waters to rivers also means that there can be a strong flush of chemicals that have built up on the surface at the start of a rainfall event. Sewage effluent is also another major pollutant. In many countries there are sophisticated wastewater treatment techniques that clean the water before it is returned to rivers. However, during major storm events many of the combined sewer systems that also take stormwater cannot cope with the amount of water entering them, and so pollution of river waters occurs as the sewage systems overflow. Even during low flow periods, sophisticated treatment may not remove a large range of compounds that have entered the water system through human use, such as antibacterial products, disinfectants, antibiotics, narcotics, fire retardants and some chemicals in soaps, shampoos and other personal care products. Research has shown that in most countries microplastics and pharmaceutical compounds are found in water bodies, often in surprisingly high concentrations. Some of these chemicals may be toxic to aquatic organisms, but more tests for impacts on aquatic ecology are required for a wide range of pharmaceutical compounds. Some compounds, such as hormones from a range of pharmaceutical products, have been found to influence the gender of fish and reduced the proportion of male fish compared to female fish. While generally not measured as a water quality variable, light pollution also impacts the ecology of freshwater systems and the interactions between water and land. Nocturnal illumination is increasing each year and artificial lights near lakes and rivers have been shown to influence insect behaviour, and thus food chain dynamics, in and around aquatic systems. Therefore, light pollution of water bodies is a growing research area, and it appears that modern LED lighting, which is very energy efficient, has stronger impacts on ecological processes than traditional lighting systems.

COASTS

Coastal areas are also subject to water pollution, much of which is derived from polluted river waters entering the coastal environment as well as pollution from ships, oil spills and some industrial effluent that is discharged into the sea. Coastal environments are important since 40 per cent of the world's population lives within 100 kilometres of the coast. Coastal environments support fisheries, large ecosystems, leisure activities and power generation, and act as an important buffer during storm events. As discussed in Chapter 4, sea-level rise is a major threat to the coastal community. According to a special IPCC report in 2019 on the ocean and crysophere, sea levels have risen by an average of 16 centimetres between 1902 and 2015 and are currently rising at 3.1–4.1 millimetres per year. Compared to the mean sea level for 1986–2005, sea levels are expected to rise by a further 43 centimetres under RCP2.6 (see Chapter 4 for an explanation of the RCP scenarios) by 2100 and by 84 centimetres under RCP8.5. Such changes will have dire consequences for coastal communities, with significantly more coastal flooding and erosion. Local high sea levels that may previously have occurred once per century will occur, on average, once per year by 2100. It is therefore important to understand the physical geography of coastal regions so that we may better anticipate and adapt to changes in the future.

WAVES

Waves are the most important feature of coastal environments as they drive many sediment transport processes, and therefore the inputs and outputs of sediment from an area. The inputs and outputs of sediment, in turn, help shape the landforms of the coastal environment. Waves are generated by wind. Stronger wind results in larger waves. Waves can travel vast distances across the oceans and so wind conditions a considerable distance from the coast can influence coastal waves. Waves can be measured for their height *H* (height between top and bottom of the wave surface), length *L* (distance between successive wave peaks (crests)) (Figure 5.8) and period (i.e. time between two wave peaks or troughs passing by a point).

Wave behaviour depends on the depth of water in the ocean. Where the water depth is more than twice the wave length, as waves travel across the water surface the water beneath moves in

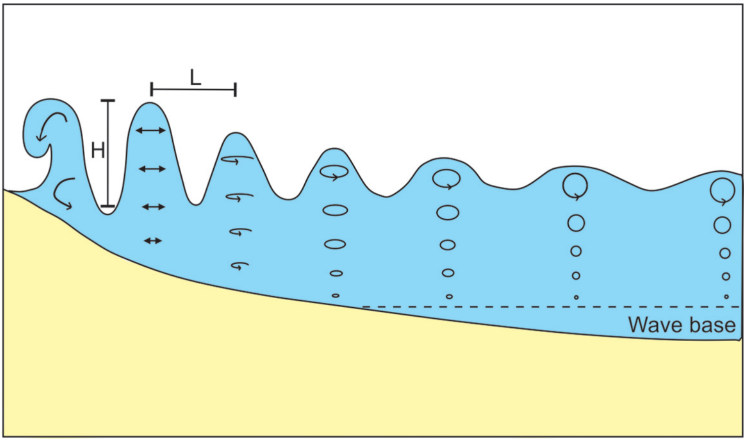

Figure 5.8 Wave height increases closer to shore as the water depth decreases. Circular water motions become more elliptical and then almost horizontal very close to the shore. Wave height (H) and length (L), indicated, are common variables measured.

circles; forward under the crest of the wave and backward under the trough. The diameter of the circles of water decreases with depth until in deep water the wave motion is no longer detectable (Figure 5.8). In water that has a more intermediate depth, where the depth is between twice the wave length and a twentieth of the wave length, the wave motion will be affected by friction on the seabed. This means that the water motion beneath the waves becomes more elliptical, with the ellipses being smaller and flatter nearer the bed, so that at the bed water just moves back and forth (Figure 5.8). In shallow water very close to the shore where the water depth is less than a twentieth of the wave length, the water movement is just horizontal in a back-and-forth direction. In reality what this all means is that as waves move from deep to shallow water they get closer together and slow down (Figure 5.8). Waves also get higher as they travel into shallower water, a process known as **shoaling**. They also change their shape from a more symmetrical wave shape to a shape with more peaked crests and flatter troughs. In deep areas, water movement produced by waves is back and forth at the same velocity (i.e. it is the shape of waves and their energy that is moving across the ocean but not the water itself within the waves). However, close to the shore the onshore side of the wave is stronger and this promotes sediment transport more favourably in the onshore direction. Wave **refraction** occurs as the wave moves close to the shore. This is when the section of the wave in shallower water travels slower due to friction on the bed than the section of the wave in deeper water. The outcome is that the wave crest rotates to become parallel with the contours of the seabed, so that wave direction 'bends' as it approaches the shore.

When the water depth is just slightly greater than the wave height then the wave breaks up, creating an area known as the **surf zone**. The energy released by breaking waves can be significant and help generate nearshore currents and sediment transport. When waves break they cause a rise in the water level on the beach (the swash), with the returning water (the backwash) running back down the slope into the sea. The swash is better at transporting sediment than the backwash, and so this helps maintain the gradient of the beach.

Nearshore currents are important for landform development in coastal areas. These currents gain their energy from wave breaking, so if the waves are stronger (e.g. during storms) then the currents will be stronger. Longshore currents flow parallel to the shore within the surf zone driven by waves entering the surf zone with their crests aligned at oblique angles to the shoreline (Figure 5.9). Longshore currents are also affected by winds and can be particularly strong when winds are blowing in the same direction as the longshore current. The net effect of the sediment movement associated with longshore currents is that large quantities of sediment will move along the coast. This is known as **longshore drift**. Another type of nearshore current, which all those who go swimming or surfing in nearshore coastal waters should be aware of, is the rip current. These are strong, narrow, seaward flowing currents that flow back through gaps between sand bars. The water piles up as a result of the waves approaching at an oblique angle and this water rushes back through the surf zone at key points.

A final important wave type is that of a **tsunami**. These are rare but can result in an enormous wave at the coastal zone, which floods coastal areas and causes massive loss of life, such as in the event on 26 December 2004 around coastlines of the Indian Ocean that killed an estimated quarter of a million people. That particular tsunami was caused by an earthquake that displaced water deep within the ocean. The tsunami killed people near the earthquake in Indonesia, Thailand and Malaysia, but also vast distances away in India, Sri Lanka, Somalia, Kenya and Tanzania. Tsunamis can also be caused by a large landslide entering the sea,

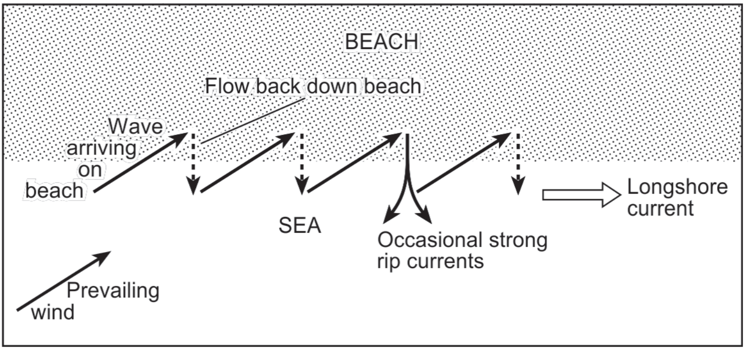

Figure 5.9 Longshore current produced by waves arriving at oblique angles to the shore.

such as that which occurred after the eruption of the Krakatoa volcano in 1883, killing 33 000 people, and the 2018 Sundra Strait tsunami, Indonesia, where over 400 people died. When a tsunami travels across the deep ocean it can typically only have a height of a few centimetres (60 centimetres after 2 hours on 26 December 2004), but they travel quickly at around 500–1000 kilometres per hour. When they reach shallower water and approach the land tsunami waves will begin to shoal, causing the water to suddenly retreat a long way from the beach, and then when the wave eventually arrives it can reach tens of metres in height.

TIDES

Tides are driven by the gravitational pull of the Moon and the Sun on the Earth. Tides have a predictable daily and monthly cycle. The Earth and Moon exert a gravitational pull on each other, which is counter-balanced by forces associated with the Earth's orbital rotation. In theory, the force exerted due to the Earth's rotation is equal on all parts of the Earth's surface, except where the gravitational pull of the Moon allows for a slight reduction. This reduction allows the oceanic surface to bulge. The oceanic bulge will occur on the part of the Earth closest to the Moon. In addition, there is a balancing bulge on the opposite side of the Earth as a product of the forces exerted by the rotation of the Earth. Consequently, there are two bulges and thus two tides per day. The tide rises and falls at a point on the Earth's surface as that point rotates away from and towards the direct line of the gravitational pull of the Moon. The gravitational pull of the Sun introduces an extra monthly dimension to the tidal sequence. As the Moon's position relative to the Sun changes over the lunar month (the Moon takes 28 days to revolve around the Earth) the two astronomical bodies either pull in alignment or opposition to each other. During the full and new moon phases they pull along the same direction, allowing the oceanic surface to bulge further and resulting in larger magnitude tides known as **spring tides**. During the half moon phases the Sun and Moon pull in opposite directions, resulting in subdued tidal bulges, known as **neap tides**. The actual impact of the gravitational pull on tides depends on the shape and topography of the coast. In some locations the tide can rise and fall by several metres, whereas in others there is a barely noticeable tide. The largest tidal ranges seem to be associated with constricted areas where there are narrow seaways connecting the sea to the main ocean, such as the Bay of Fundy, Canada (17-metre tidal range) or the Irish Sea (13-metre tidal range in the Severn Estuary), or where there are wide continental shelves, such as off the east coast of China. On coasts facing the wide, open ocean, tidal ranges are usually less than 2 metres.

As the tide rises and falls against the coastline it produces a flow of water. This is a **tidal current**. If you throw a stick into the water at the shoreline during a rising tide the stick is pushed up the beach. However, if you throw it in at the same point on a falling tide the stick will often float out to sea as the tidal current takes it away from the shore. The geometry of the coastline can control the tidal current, which is most pronounced in river mouths, estuaries and where flow is squeezed through inlets. A **tidal bore** can occur in a few such locations on Earth, such as the Bay of Fundy or the Qiantang River in China, whereby the leading edge of the tide creates a counter current along shallow river mouths that is temporarily stronger than the riverflow from land, with a resulting wave crest appearing that travels upriver.

During strong storms the coastal water level can be higher than the normal tidal level. This elevated water level above the predicted tidal level is known as a storm surge. Such surges can be several metres high during cyclones, with their height depending on atmospheric pressure, the onshore wind and the coastal geometry. Sea level is raised by 1 centimetre for every millibar fall in atmospheric pressure, and storms are associated with low pressure. If the wind blows towards the shore water levels can be forced higher against the coastline. Shallow gradient coasts with funnel shaped inlets can be extremely prone to storm surges, such as the North Sea or the Bay of Bengal. Storm surges can cause widespread damage and severe flooding of coastal areas. Sea-level rise combined with increased storm intensities under future climate change means that storm surges may become more common and severe in the future.

COASTAL LANDFORMS

The factors that help shape coastal landforms include waves, tides and rivers, with the nature of the land material at the coast also being an important factor (e.g. hard rock or erodible loose sediment). There are different landforms that are characteristic of locations where waves, tides or riverine processes dominate. For example, a common feature of wave-dominated coasts is a beach, perhaps with coastal dunes.

Beaches are formed by deposits of sediment brought by waves and are typically shaped with a concave profile. Towards the top of the beach is the **berm**, which is where the slope steepens and then flattens off. Within the beach there can often be cusps of sand or gravel at the shoreline, which are repetitive features a few metres apart formed by swash action. Beaches respond to changing wave energy conditions. When it is calm and wave energy is low, the net sediment transport is in an onshore direction, resulting in steepening of the beach and a pronounced berm. In stormy conditions, net offshore sediment transport occurs with the destruction of the berm and flattening of the beach. This then helps to dissipate the wave energy over a wider area.

Coastal dunes protect the coastal area behind them by providing a buffer to extreme waves and winds. Dune formation requires wind and a large supply of sand (see also Chapter 2). Onshore winds capable of sediment transport must occur for a significant amount of time. Dunes often develop just above the spring high tide line where litter such as seaweed and wood collect. This litter starts to trap sand that is blowing around and promotes the formation of small dunes. Once these proto-dunes start to form they can then enlarge, especially when plants grow within them, allowing a greater height of sand to be trapped around the plants. Dunes can grow relatively quickly under the right conditions, reaching 2 metres height after just 5 years (see also erosion section in Chapter 2).

Just off the coast in wave-dominated environments there can be formation of barrier landforms, including barrier islands and lagoons. These extremely dynamic landforms can be found along about an eighth of the world's coastline and there are many famous examples, such as those along the Florida Atlantic coast, the coast of The Netherlands, many Arctic coastlines and off the coast of North Carolina (Figure 5.10). Barriers help buffer inland areas from storm wave energy and represent a large accumulation of onshore moving sand bars. As can be seen in Figure 5.10, barriers

Figure 5.10 Barrier islands around Cape Hatteras, North Carolina. The sheltered lagoon can be seen between the mainland and the barrier islands. The image represents a distance of 70 kilometres in width. (Source: NASA's Earth Observatory, Jesse Allen, https://earthobservatory.nasa.gov/images/88619/cape-hatteras-national-seashore.)

form long strings of parallel island chains punctuated by tidal inlets that allow the transfer of water and sediments between the open sea and the lagoons behind the barrier. Some barriers are aligned to the swash direction, whereas others are aligned to the prevailing direction of longshore currents. For example, a **spit** is a narrow accumulation of sand or gravel with one end attached to the mainland and the other projecting into the sea or across the mouth of an estuary or a bay. Spits grow in the longshore drift direction and can only exist where there is a regular supply of sediment.

Estuaries are river mouth locations where sedimentary deposits from both river and sea sources create landforms. Sea levels rose around river mouths as ice melted following the last glacial period, stabilising approximately 6 000 years ago (although sea level is increasing again with contemporary climate change). Infilling of estuaries occurred as sediments from land and sea could inundate the new deep waters at the river mouth. Estuaries have three broad zones; an upper, middle and outer part. River processes dominate the upper estuary, while marine processes dominate the outer estuary. The middle section is mixed. The same volume of sea water leaves the estuary during the falling tide as enters during the rising tide. However, the duration and strength of the rising and falling tide tends to be different. Estuaries or channel sections that display a flooding tide that is faster and stronger than the ebbing tide are said to be flood-dominant, whereas those that display an ebbing tide that is greatest are ebb-dominant. Flood- or ebb-dominance influences whether net sediment transport is landward or seaward, respectively. Flood-dominant estuaries infill their entrance channels by continually pushing coastal sediment landward, often causing the exit to become clogged up and meaning the estuary mouth bathymetry regularly changes, which makes their use for shipping difficult. Ebb-dominant estuaries tend to flush sediment seawards and are more stable environments for shipping.

The outer zone in most estuaries is devoid of vegetation due to excessive sediment and water movement. Further within the estuary, however, salt tolerant plants can grow. In tropical environments mangrove systems may be prevalent and in temperate environments there are salt marshes. These ecosystems significantly enhance sediment deposition. The steady supply of organic matter from the plants also adds to the sediment, which means that these zones rise in altitude over time as the sediment accumulates. This build-up of sediment is often at a faster rate than sea-level rise, meaning that these environments can keep up with sea level as long as they are protected from human damage. Such effects have been recently observed on the southwestern coast of Florida where sea-level rise is driving enhanced growth of mangroves, with additional carbon accumulation in the sediment as the system grows in altitude over time to keep up with sea levels.

Deltas are coastal landforms dominated by river processes. They are accumulations of sediment deposited where rivers enter into the sea. Here, the amount of sediment delivered by the river is greater than that removed by waves and tides, and so the delta moves seawards. Although there are not many deltas around the world, as they tend only to be associated with large river systems (e.g. Mississippi, Ganges, Lena, Nile), large populations of people live on deltas. The coarsest sediments are deposited close to the river mouth and the finer sediments settle out further seaward. Ongoing delta survival and development relies on an active sediment supply by the river. Deltas can be very dynamic. As one area of the delta grows upwards due to building sediment, the river channels can suddenly no longer keep flowing to that raised area and so they move across the delta, supplying sediment elsewhere. The starving of sediment from some areas of delta can then lead to net erosion through wave and tide action. This dynamism creates a hazard for humans living on deltas.

On the other side of the spectrum, rocky coasts may be seen to be solid, more stable landforms. However, they are actually characterised by erosional features. This often makes for stunning scenery with rugged cliffs and interesting stack features, arches and caves. Nevertheless, the average rate of change along rocky coasts is slow, although some changes, such as landslips, can be dramatic. Rocky coast erosion occurs through mass movements, rock weathering processes and rock transport processes. Mass movements (see Chapter 2) are common due to steep slopes. Rockfalls are characteristic of hard rocks, landslides typically occur in thick deposits of clay, shale or marl, and flows occur when there is a high liquid content. Freeze-thaw, wetting and drying, chemical weathering, and the mechanical abrasion and force of water undercutting the base of rocky slopes by wave processes are principal drivers of mass movement. These processes are discussed in Chapter 2. Typically, once material has been removed from the cliff face onto the floor below it can be removed by coastal transport processes, meaning that the base of cliffs do not receive protection for long by fallen debris. **Shore platforms** develop when erosion of a rocky coast leaves behind a horizontal or gently sloping rock surface.

Coral reefs are depositional environments yet are located in high energy wave systems. They are the largest biologically constructed formations on Earth and consist of limestone that has been created by animals forming their shells. When corals die, they leave behind the limestone that formed their skeletons, and so the sediment can build up over thousands of years. Coral reefs tend to exist in a very delicate balance between erosion and biological construction. Without the presence of living organisms the reefs would not exist because weathering and wave erosion would destroy the landforms. Corals can be found throughout the world, but reef building corals are only found in the sub-tropics between 30°N and 30°S.

Coral reefs occur in two main settings. The first is close to land on the continental shelf where water depths are less than 200 metres, such as in the Great Barrier Reef off the north-east coast of Queensland, Australia. The second are those that rise several kilometres from the ocean floor. These have formed around the edges of volcanic islands above hot spots (see Chapter 2 for an explanation of hot spot oceanic island formation). Often when the volcano becomes extinct, the island erodes and sinks back into the ocean. The sinking occurs because the oceanic crust on which the volcano rests cools and also because the whole crust sinks as it moves away from the mid ocean ridge towards the subduction zone where it meets continental crust (see Figure 2.1 in Chapter 2). However, coral growth may be able to keep up with the rate of island sinking and so there may be a large depth of limestone formed on top of the volcanic base. The key is whether vertical reef growth can keep pace with falling land levels or rising sea levels. Atolls are reefs that surround a central lagoon and most of these are found in the Indian and Pacific Oceans. They tend to be circular in shape and range from 75 kilometres width to less than 1 kilometre. Many are believed to have developed around the rim of a volcanic island. As the volcano has itself sunk back into the ocean, the coral has been able to keep up growth, although only around the edges where the coral had previously developed. Thus, the central portion of the island is largely devoid of reef and a lagoon forms. The nutrients in the lagoon are poor as wave action is restricted, and thus growth is limited.

COASTAL MANAGEMENT

In addition to sea-level rise, coastal managers have to deal with natural erosion and depositional processes and those processes that have been created by human action (e.g. beach erosion due to sand abstraction somewhere further along the coast, which starves longshore drift of sediment). Many coastal management solutions have involved hard engineering, such as sea walls, breakwaters and groynes. Sea walls are large and costly and usually made from concrete, steel or timber, often with a curved face. However, as sea walls provide a limit to the beach zone, they may stop the usual process of beach profiles changing (lengthening/flattening) during stormy periods and steepening during calm periods. Sea walls also reflect waves, so instead of dissipating energy more erosion takes place at points further along the coastline, beyond the extent of the wall. A common response is to then extend the seawall further, however this just shifts the erosion problem further along the coast. A sea wall was built in the nineteenth century at the seaside resort of Blackpool, England, and 150 years later almost 80 kilometres of coastline in the area had a sea wall installed as erosion developed along the coast from the Blackpool wall.

An alternative strategy to sea walls is to reduce the incoming wave energy by installing submerged breakwaters parallel to the shore that encourage waves to break further away from the land. These features must be porous, so that they allow sediment to move through them, and are usually built as a series to protect a stretch of coast. They are very expensive to install because they must withstand extreme wave action and are placed within the most energetic part of the nearshore zone. Building beaches is another solution to coastal erosion. This involves artificial deposition of sediment on the beach or in the nearshore zone in order to advance the shoreline seaward. However, this sort of approach can often treat the symptoms and not the causes of the erosion problem and beach nourishment can be followed by net erosion. Officials in Miami Beach, Florida, spend millions of dollars every few years transporting sand to replenish the beach. Generally, the sand used for beach nourishment must be coarser than local sediment in order to minimise rapid sediment loss offshore. **Groynes** are often installed in beaches to trap sediment moving via longshore drift. The groynes can, however, be buried by sediment, allowing longshore drift to recommence. A problem with groynes is that rip currents sometimes develop on their downdrift side, which consequently moves sediment offshore and away from the beach system. Therefore, some groynes are constructed with bends to counteract this effect. Jetties are built to line the banks of tidal inlets or river outlets to stabilise the waterway for navigation. However, these jetties often encourage deposition on the updrift side and erosion downdrift. This occurred in the 1920s at Santa Barbara, USA, where the harbour infilled with sediment. Further down the coast, as a result, the whole community became at risk from coastal erosion.

Overall, there are probably four main strategies for dealing with coastal erosion and sea-level rise. These are: doing nothing, abandonment, adaptation or protection. The first option might be most costly as it leaves infrastructure and people at risk, and so can only be realistic if the area is sparsely populated. Abandonment is where people and industry leave the coastal setting and move inland, and also where further coastal development is prevented. Adaptation involves designing features into the landscape to cope with change, such as building homes on stilts or providing warning systems. Finally, protection involves engineering solutions, such as tidal barriers with large sluice gates on estuaries to prevent very high tides from entering the estuary. The challenges for coastal management are enormous because the system is dynamic, requiring flows of sediment, water and energy. Interfering with those flows in one place has knock-on effects further along the coast. Some coastal areas are currently responding most to short-term changes caused by human action (e.g. see Box 5.5), whereas others are still responding to changes in sea level since the retreat of ice sheets began at the end of the last glacial period some 18 000 years ago.

BOX 5.5 MANAGING COASTAL RETREAT IN WEST AFRICA

Along the West African coast, retreat is a major threat, occurring at rates of between 1 and 2 metres per year. In some places where mangroves have been removed by humans, coastal retreat can occur at tens of metres per year. The mangroves are often removed for firewood and to clear land for farming. However, the coastal erosion can be devastating to local economies, and in West Africa 40 per cent of gross domestic product is related to coastal activity, including cities, ports, fisheries, petroleum production and food processing. Tackling coastal erosion in the region has seen investment in hard engineering, with only 3 per cent going towards nature-based solutions. However, recent recommendations have focussed on the importance of protecting natural systems such as mangroves, which provide vital protection to coastlines by trapping sediment, raising the land surface and buffering against storms. Thus, the establishment of the West Africa Coastal Areas Management Programme has been an important springboard to delivering shared activity in West African coastal regions. Mangrove protection and planting has begun and general work is underway towards a better understanding of the risks in order to inform policies and plans and design cost-effective risk-reduction solutions. There is also work on combining hard infrastructure with ecosystem-based solutions (such as mangrove preservation) and proper land-use planning, moving from a purely ‘holding the line’ strategy to consideration of accommodation and managed retreat.

For more information please see: www.wacaprogram.org.

ICE

Ice exists on Earth today mainly in large ice sheets over the Antarctic and Greenland, in ice caps, in sea-ice over the North Pole and in valley glaciers on land. It also exists in large quantities within frozen ground known as permafrost and produces characteristic landforms in cold regions. If the amount of ice present today in the Greenland ice sheet melted this would raise sea levels around the world by 7 metres. Ice sheets form over very large areas, the size of continents, and are typically a few kilometres thick. Ice sheets flow slowly, although there can be faster moving rivers of ice within them. In Antarctica the faster moving ice streams feed ice shelves, which are formations of floating ice that then melt into the ocean or are broken away and float off into the ocean.

GLACIER AND ICE SHEET DYNAMICS

Glaciers are much smaller than ice sheets, covering individual valleys on land, but they occur on every continent. Glaciers are formed in mountainous regions wherever snow accumulates at a faster rate than it can be melted. This therefore requires both a good supply of precipitation and cold conditions. A typical glacier has an accumulation zone at the top where rates of gain are more than losses and an **ablation zone** at the bottom where rates of loss are greater than rates of gain. The loss of ice at the outlet of the glacier mainly occurs through melt water but sometimes, if the glacier directly flows into the sea, icebergs can be broken off and float away before melting.

As the melting point of ice reduces with increasing pressure there is an important difference to be made between 'warm' ice and 'cold' ice. At a depth of 1 kilometre within an ice sheet the melting point of ice occurs at the cooler temperature of −0.7°C, rather than 0°C. Therefore, the thicker the ice mass the more likely it is to produce water at depth at the **pressure-melting point**. Warm ice is at the pressure melting point and contains liquid water, whereas cold ice occurs at temperatures below the pressure melting point and does not contain liquid water. The pressure melting point concept is important for understanding what might be happening within the ice and at the base of the ice mass because water can help glaciers move over their bed. Warm ice occurs throughout temperate glaciers except near the surface of the glacier, which becomes cold in winter. Cold ice occurs throughout cold glaciers. If a glacier bed is cold, flowing water will not occur and so there will be less sliding and deformation of sediments than under a warm ice base. In some places the heat released from the Earth, produced by tectonic activity, can melt ice at the base of an ice mass. There are more than 400 lakes below the Antarctic ice mass, the largest being 4 kilometres below the ice surface and covering around 14 000 square kilometres. Many of these have been isolated from the atmosphere for millions of years. One of the theories behind why these large lakes exist is that they are warmed by geothermal and tectonic activity occurring below them. However, little is known about these lakes and they are difficult to study because of their location. Whether these lakes contain ancient forms of living microbes remains to be discovered, but drilling to date has led to the discovery of crustacean remains showing previous life in one of these lakes.

Water is produced at the snow surface of many glaciers in summer. Unless it refreezes, meltwater may percolate downwards through any snow and run across the surface of the glacier ice. Water will flow downslope along the ice surface and will emerge as a series of small streams, many of which will have created smooth channels on the ice surface. The water may flow to the outlet of the glacier or some of it may flow within the glacier in tunnels. Water that descends from the glacier surface may arrive at the bed at discrete locations flowing in channels. Two types of channels exist beneath glaciers: Nye (N) channels and Röthlisberger (R) channels. N-channels are formed in the bedrock and R-channels are formed upwards into the ice.

To move ice from the accumulation zone downslope to the ablation zone the glacier must physically flow downslope. There are three mechanisms by which glaciers flow: internal deformation or creep (sometimes called plastic flow), sliding (sometimes called basal slip) and bed deformation. Stress applied under the action of gravity causes the ice to deform and to creep along. While the creep rate is much lower for cold ice than warm ice, creep is, overall, a slow process. Once the pressure melting point is reached at the glacier bed, sliding can occur because the presence of water reduces the friction. If the bed of a glacier is cold, the sliding is restricted. Glacier bases often contain rock obstacles. Here **regelation** can occur. As ice flows around obstacles at the bed there will be excess pressure upstream of the obstacle and lowered pressure on the downstream side. Increased pressure lowers the ice melting point upstream of the obstacle. The melted water then flows around the obstacle to the low-pressure downstream side where it refreezes because the melting point is higher. This mechanism therefore allows the ice to slide past the obstacle. Regelation is limited by heat conduction, which is better for small obstacles. The regelation process is important as it often causes obstacles to get frozen into the base of the ice and then moved with the ice mass. Where an ice sheet or glacier sits on soft sediments, movement of these sediments can also assist ice movement as the sediments themselves deform. This is bed deformation.

The rate of movement of glaciers and ice sheets is very variable (Box 5.6). Temperate glaciers often flow at tens of metres per year, whereas cold glaciers flow at rates of 2 metres per year or less. Some fast-flowing ice streams can move at several hundred metres per year by fast sliding or on sediment that is deforming. Around 1 per cent of glaciers experience sudden phases of surging and then quiet, slow-moving periods. For example, on Svalbard in the Arctic, the slow periods can last for around a century, while fast moving periods may last for 1–5 years. In 1953, the Kutiah Glacier in Pakistan advanced 12 kilometres in only 3 months. It is uncertain why this surging occurs but a combination of factors is likely, including a build-up of meltwater which can trickle down and accumulate at the bed.

BOX 5.6 SATELLITE OBSERVATIONS OF ICE SHEET LOSS

Two key land stores of ice are thought to be major potential contributors to sea-level rise: the Antarctic and Greenland ice sheets. These systems are so vast it is difficult to survey the mass of ice and how it is changing based on data from ground observations. Therefore, we rely on satellite observations of ice elevation change, motion and thickness to provide key data on ice sheet changes. A recent study published in *Nature* by the IMBIE team of scientists (Shepherd et al., 2020) used 26 different satellite measurements of changes in the Greenland ice sheet altitude, velocity and gravitational potential to produce a combined estimate of its mass balance. They showed gravitational potential to produce a combined estimate of its mass balance. They showed that the Greenland ice sheet was in an approximate state of balance in the early 1990s (inputs from precipitation approximately equalled outputs from melting) but that net annual losses have risen since then, peaking at 345 ± 66 billion tonnes per year in 2011. The losses since 1992 caused the mean global sea level to rise by 10.8 ± 0.9 millimetres, making it the largest single contributor to sea-level rise. The rate of ice loss varied from 26 ± 27 gigatonnes per year between 1992 and 1997, peaking at 275 ± 28 gigatonnes per year between 2007 and 2012, and reducing to 244 ± 28 gigatonnes per year between 2012 and 2017. Around half of the ice losses were due to enhanced melt associated with warmer atmospheric and oceanic conditions and the other half appears to be attributable to faster flow of some of the major outlet glaciers on the ice sheet.

A follow-up study by King et al. (2020) used satellite data measuring outlet glacier velocity, elevation and front position changes over the full ice sheet. Changes are not evenly distributed across the ice sheet (Figure 5.11). They found that increased glacier discharge was due almost entirely to the retreat of glacier fronts, rather than inland ice sheet processes. This widespread retreat between 2000 and 2005 resulted in a big increase in discharge and a switch to a new dynamic state of sustained mass loss, which has continued ever since.

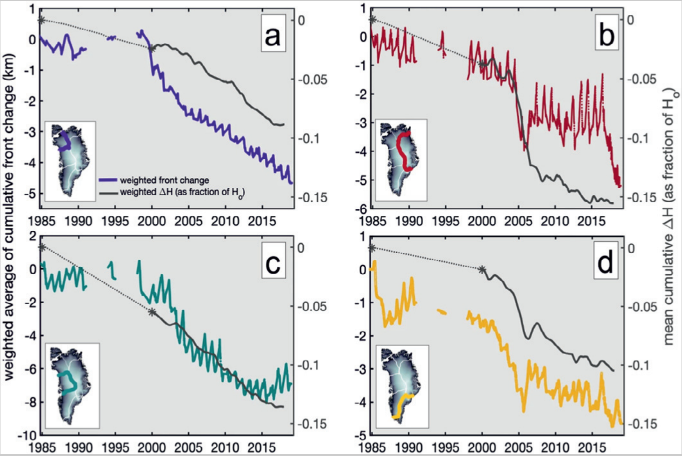

Figure 5.11 Net regional trends in ice thickness (black curves) and ice front position (coloured curves) in a) northwest, b) central east, c) central west and d) southeast parts of the Greenland ice sheet, with the colours corresponding to the locations mapped. Negative values on the left y-axis indicate net cumulative retreat over the study period. Lighter, dotted grey lines between the two starred points indicate linear interpolation between two sparse data points, and the apparent trend during this period should be treated with caution.

(Source: King et al., 2020. CC BY 4.0: http://creativecommons.org/licenses/by/4.0/.)

For more information about the studies mentioned in this box see the original research papers cited. Please also visit the IMBIE website: http://imbie.org and the Centre for Polar Observation and Monitoring website for the latest information: https://cpom.org.uk

GLACIAL LANDFORMS

Glacial erosion removes enormous volumes of rock and produces characteristic landforms. The mass of ice can crush rock where there are weaknesses, producing angular sediments. Crushed bedrock can be taken up into the glacier. The process by which a glacier removes large chunks of rock from its bed is known as plucking. Plucking can occur by ice regelation around the rock or by incorporation into the ice along faults. Glacial abrasion occurs when rocks and particles at the base of the ice slide over the bedrock, thereby scratching it and wearing it down. The sediment resulting from abrasion is very fine and when suspended in water is known as glacial flour. Meltwater at the bed of a glacier can also cause erosion by mechanical or chemical processes, as with a normal river system. The rate at which glaciers and ice sheets erode varies depending on temperature, the rate of ice movement and the bedrock. On average, across the planet, it seems that a rate of 1 metre of erosion for every 1000 years is a reasonable estimate for ice masses.

Weathering action such as freeze-thaw and undercutting by the glacier results in material falling from rock slopes onto the surface of glaciers, where it is transported downstream by the ice mass. If the material falls onto the glacier in the upper accumulation zone it will become buried within the glacier. The material will be transported either within the main volume of ice or it may slowly descend to the bed, eventually assisting with the scouring of the bedrock. If material falls onto the surface in the ablation zone it will usually remain on top of the glacier. There have been some interesting materials found incorporated and preserved by glacier ice. For example, in Siberia, an extinct woolly mammoth was found frozen inside a glacier, while in 1991 in the Alps a pre-historic human was found, nicknamed Ötzi.

The features formed by glacial erosion include large-scale landforms, such as U-shaped valleys with steep ridges and horns between the valleys. Here, ice masses flowed over the landscape, scouring it as they moved and leaving angular summits exposed that protruded above the ice. Glacial U-shaped valleys or troughs (note that river-cut valleys are usually V-shaped) are mainly the product of abrasion, but rock fracturing and plucking downstream of smoothed obstacles are also important. While most glaciers follow existing river valleys, they act to deepen, widen and straighten them. Lakes often form in the eroded valley left by a glacier and these lakes slowly infill with sediment over time. If the glacier has eroded the valley to below sea level, then the valley may get flooded after the ice has melted, forming a fjord. Some infilling of the valley bottom occurs after ice melt as the gravels produced by meltwaters downstream of the retreating glacier, combined with lake sediments, act to flatten the floor of the valley, assisting the production of the U-shape.

Often side valleys to the main glacial valley are truncated, leaving **hanging valleys** (Figure 5.12a). These hanging valleys were occupied by tributary glaciers feeding into the main glacier. The landscape of Yosemite in California is a good example where there are plentiful hanging valleys, often with large waterfalls tumbling over them for several hundred metres down vertical cliffs into the main valley. Small glaciers often form dome-shaped depressions, known as **corries** or **cirques**, near the top of mountains. Once the ice has retreated the depression can form a small lake, often called a **tarn**. The cirques are essentially small forms of hanging valleys. Glacial valleys and cirques may extend backwards, particularly as freeze-thaw activity is intense, resulting in rock shattering. Two cirques that cut backwards into each other may leave a narrow ridge between them known as an **arête**. If three or more cirques cut back together, they may leave a pyramidal peak or horn, such as Mont Blanc in eastern France. Sometimes, weathering or erosion of part of an arête can produce a dip in it known as a **col**. These often form low points in mountain ridges that humans have turned into routes to get across the mountains, as is the case for many Alpine passes.

On a smaller scale there are smoothed **whaleback forms**, ranging from 10 metres to hundreds of metres in length, which are

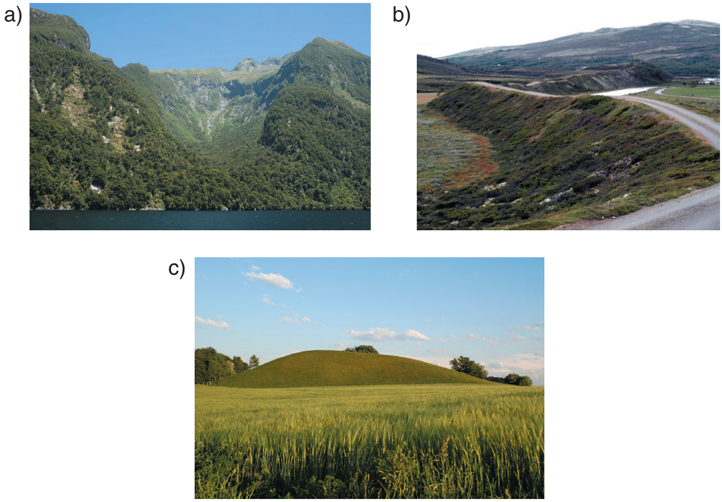

Figure 5.12 Example features of formerly glaciated landscapes: a) hanging valley above Doubtful Sound fjord, New Zealand; b) esker at Einunndalsranden nature reserve, Hedmark, Norway; c) grassed over drumlin in a field, Andechs, Germany.

(Source: a) Kaitil; b) Boschfoto; c) Pseudopanax; a) and b) CC BY-SA 3.0: https://creativecommons.org/licenses/by-sa/3.0.)

the product of abrasion across the surface of an obstruction. The rock may have been particularly resistant, meaning the glacier was unable to abrade it fully, thereby leaving a smoothed mound orientated in the direction of glacier flow. However, many of these small mounds that are the product of erosion are not smooth all over. **Stoss-and-lee forms** are streamlined features that have a gently sloped, glacially smoothed upstream side and a steeper, plucked, downstream side. They are much more common than whaleback forms and occur when the glacier flows over the obstacle, smoothing the upslope side. The stoss side is often scratched with grooves (**striations**) from sediment that has abraded the rock as it passed over it. On the downslope side of these forms, bedrock fracturing, loosening and displacement can occur and the fragments can be incorporated into the ice. Therefore, the downslope side of stoss-and-lee forms tends to be rough. Small stoss-and-lee forms are often called **roche mountonées**. These landforms tell us that there was warm ice in the glacier that once existed at that location. Roche mountonées are typically a few metres in height and tens of metres in length. There are many remarkable undulating landscapes where there are hundreds of roche mountonées spread over large areas. **Crag and tail features** form where resistant rock leaves small mounds protruding above the surrounding surface, with sediment deposited on the downslope side of the feature, behind the obstruction. The setting for Edinburgh Castle, in Scotland, is a crag and tail feature.

The deposition of eroded material from glaciation has also produced important landforms. The glacier itself can produce the landform, or the landform may be left when the ice in the glacier melts. In addition, meltwater can produce depositional features quite some distance from the glacier itself. Depositional landforms associated with glaciers are not as dramatic as those produced by erosion. Nevertheless, glacial deposits cover around 75 per cent of the mid-latitude land mass and 8 per cent of the Earth's surface. Features may be formed by the direct action of ice, such as **moraines**, which are linear mounds of sediment. Push moraines form when a glacier forces itself into sediment at the front of the glacier, raising it into a ridge. These moraines can mark the maximum extent of glaciers in the past, helping us to map the extent of former ice masses. Dump moraines are formed at the front of a glacier where material being transported through the glacier is eventually deposited at its front end as the ice melts. There can be many of these left as a glacier retreats. Lateral moraines are those that have formed on the glacier surface, collecting rockfall from the cliffs above. As the glacier moves, the intermittent rockfall debris appears as a linear feature on the glacier surface. Massive boulders can be transported in this way, and once they are deposited a long distance from their original source they are often known as **erratics**. Hummocky moraines occur where there is a deposit of material from inside the glacier or on top of the glacier as the glacier melts.

Moraines are formed by the action of ice, whereas **eskers** are features of water. Eskers are snaking ridges of sand and gravel (Figure 5.12b) that are thought to form mainly in R-channels at the glacier bed. They can be 20 to 30 metres high and up to 500 kilometres in length. Eskers may flow uphill as well as down, simply representing the fingerprint of an internal channel system within the ice mass. **Drumlins** are streamlined mounds of sediment, sometimes with a rock core, aligned in the direction of ice flow and typically with a blunt end upstream and a more pointed end downstream (Figure 5.12c). They often occur in large numbers across an area, such as the 10 000 or so that occur in west central New York State. There is considerable debate as to the mechanisms that cause drumlin formation, with some suggesting they were formed during really large flood events.

The outwash plain downstream of a glacier is often rich in gravels and boulders, and the meltwater system also contains a lot of rock flour and has a milky colour. Braided river systems are common in these environments. When large blocks of ice get swept down the river system and are deposited and buried, they can later form **kettle holes**. The slowly melting ice leaves a depression in the surface, where elsewhere in the surrounding landscape sediment had been deposited by the river system. On the other hand, braided rivers may deposit sediments that are several hundred metres thick in some locations, masking underlying features.

PERMAFROST

Permafrost refers to soil or bedrock that is frozen for longer than 2 years. It covers around 25 per cent of the land surface, although this is mainly in the northern hemisphere where there are significant land masses at high latitudes. More than half of the land area of both Russia and Canada, the world's two largest countries, is covered with permafrost. Permafrost usually consists of soil that contains ice within its pore spaces. Most of the Earth's permafrost is very susceptible to climate change because it exists at temperatures just a few degrees below 0°C.

At high latitudes, but outside of ice sheets and glaciers, permafrost tends to be continuous and often extends several hundred metres in depth. At slightly lower latitudes the permafrost may not be continuous and is also thinner, perhaps just being a few metres in depth. Continuous permafrost moves to discontinuous permafrost at a mean annual air temperature of around −6 to −8°C, while discontinuous permafrost extends to a mean annual temperature of around −1°C within continental interiors. Close to the surface there is an annual melt and freeze cycle in both continuous and discontinuous permafrost. Temperatures measured in permafrost are typically lowest close to the ground surface, excluding the **active layer**, and increase with depth, reaching the melting point at the base of the permafrost.

The near-surface layer in which thawing and freezing occurs is called the active layer. This layer can range from just a few centimetres in continuous permafrost to several metres deep in discontinuous permafrost zones. Freezing of water causes an expansion of 9 per cent in volume and so freeze-thaw processes in the active layer cause movements of the ground surface. In fact, this expansion and contraction can lead to major problems for infrastructure such as roads, pipelines and buildings, which, if not properly engineered to cope with the conditions, will collapse, buckle and subside. Human-occupied buildings transmit heat to the ground and can melt permafrost, leading to subsidence. Therefore, buildings must be well insulated at ground level or even have refrigeration units to cool the ground, although this can be prohibitively expensive. Ensuring that buildings are anchored to deep piles that are embedded to the bedrock is important and means that the building could be supported off the ground if the ground subsides. Unfortunately, due to climate change, large areas of permafrost are melting, and as a result some infrastructure is becoming damaged. Winter temperatures in Alaska and western Canada, for example, have increased by 3°C over the last 50 years, and climate projections suggest that for a quarter of the region the top 2–3 metres of permafrost will melt by 2100. As a result, foundations will sink. Therefore, even more robust piling and foundations are being constructed. The trans-Alaskan oil pipeline is a classic example of engineering that takes account of permafrost. The pipeline carries oil at 65°C across 1 285 kilometres of permafrost terrain. The hot temperature of the pipeline would thaw the ground if the pipe were buried or at the ground surface, which would result in subsidence and damage to the pipeline. Therefore, it is elevated above the ground on racks for a large proportion of its route. The pipeline was built with bends in it to allow the pipeline to expand and contract, and move sideways and vertically without cracking. The vertical supports for the pipe are also equipped with devices to cool the permafrost.

LANDFORMS IN PERIGLACIAL REGIONS

Periglacial environments are those that are cold and subject to intense frost action but which are non-glacial. Permafrost regions are periglacial but permafrost is not a pre-requisite for a periglacial region. Many periglacial landforms are related to the freezing of water in soils and sediments. Frost heaving and thrusting are the vertical and horizontal movement of sediment due to the formation of ice. Heaving generally dominates because ice crystallisation tends to happen in a direction parallel to the temperature gradient, and in the soil this is usually parallel to the ground surface. Frost heaving moves masses of soil and may even push stones upwards to the surface. This may happen as the stone and surrounding soil is pushed upwards during freezing in the active layer. When melting occurs in the summer, the finer sediments settle back down, filling in the gap below the stone and supporting it. Additionally, soil water flowing around the stone may flow into the pore spaces below it and, when this water freezes, push the stone up once more. This upwards motion leaves the stone a little higher in the soil profile each year. Over long periods, the result is a net movement of stones to the surface.

The types of mass movement described in the section on weathering and erosion in Chapter 2 occur in periglacial areas. However, **frost creep** and **solifluction** are more important in periglacial environments. Frost creep occurs when sediment is pushed upwards on a slope during freezing as part of frost heave, but then gravity pulls the sediment in a downslope direction when it melts and lowers, so that over many years there is a net movement downslope. Frost creep will operate in conjunction with solifluction. Solifluction is the slow downslope movement of saturated soil in a very slow flow. In periglacial areas the process often occurs in the active layer above permanently frozen ground. A more rapid movement of saturated soil can occur in this active layer than that which occurs in more temperate or tropical zones. The solifluction process in periglacial areas is known as gelifluction.

Water can either freeze within pore spaces between solid particles of soil or sediment, or it can migrate to form discrete masses of ice known as **segregated ice**. Coarse gravels and sands are highly permeable but because pore spaces are large there is little ‘suction’ potential (see earlier in this chapter on soil hydrology) and so they do not retain much water. Finer soils, such as clay, have low permeability but high water-retention capacity. This means that soils with intermediate grain sizes, such as silt, have the greatest potential to form segregated ice within the ground. Segregated ice may form lenses or form bands. Where the bands are thick, sometimes up to several metres, they are known as **massive ice**. A large body of experimental research has shown that liquid films can coat ice surfaces even when the temperature is below the pressure-melting point. These films provide fluid conduits that supply the growth of segregated ice. Hence there is slow movement of very thin films (a few hundreds of a thousandth of a metre in thickness) of water that coats the ice and enlarges it. The movement of these thin films occurs from the sediment adjacent to the ice through several, quite complex, mechanisms, including molecular attraction, which is more powerful than the forces holding the water within the pore spaces. This then leaves relatively empty pore spaces near the segregated ice that become filled during the summer by meltwater seeping into the soil from above through suction processes. This means the segregated ice can slowly grow.

Frost cracking can occur by fracturing the ground as it contracts at very cold temperatures. While there is expansion when water freezes, as the temperature gets very low, ice and sediment contract in volume. This is probably a major factor in the formation of many polygonal crack features seen in the periglacial landscape. Frost crack polygons are usually 5 to 30 metres across and develop best when there is no insulating snow cover. These features cover large areas in North America and Siberia (e.g. Figure 5.13a). Water entering the crack can later freeze and then these ice films expand over several years of seasonal melting and freezing to become **ice wedges**. These ground ice features can be 4 metres deep and 2 metres wide, forming a V-shaped ice wedge in the enlarged crack. Therefore, vertical ice wedges in the sediment often accompany patterned frost crack polygons on the surface.

In addition to crack polygons there are often regular geometric patterns of stones or topography in periglacial areas (e.g. Figure 5.13b). These can be grouped into circles, nets, polygons, steps and stripes. These are amazing features in the landscape and almost look like humans have made patterns by sorting stones and vegetation into neat

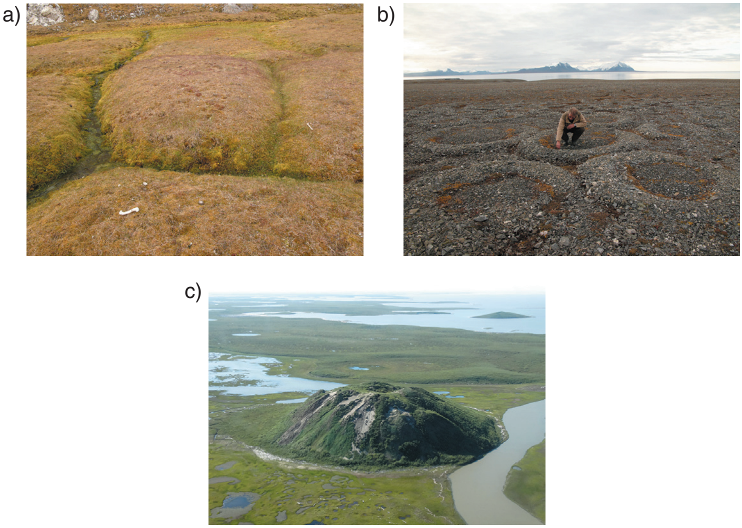

Figure 5.13 Classic periglacial landforms: a) ice wedge polygons; b) sorted stone circles; c) pingo, near Tuktoyaktuk, Northwest Territories, Canada.

(Source: a) and b) Hannes Grobe CC BY-SA 2.5: https://creativecommons.org/licenses/by-sa/2.5/deed.en; c) Adam Jones CC BY-2.0: https://creativecommons.org/licenses/by/2.0/deed.en.)

shapes. Circles, nets and polygons are common on flat surfaces, whereas steps and stripes occur on slopes between 5° and 30°. On steeper slopes, mass movement becomes a more dominant process, destroying patterned features. Sorted stone circles typically have fine material in the centre of an area of lowered relief, with larger sediment forming a higher perimeter. Sorted stripes look rather like a recently ploughed field, with ridges and furrows consisting of alternating stripes of coarse and finer material. There are several mechanisms hypothesised for these features linked to heave and thrust processes, and the possibility of mini convection circulation cells operating within the sediment near the surface. In the daytime, especially in summer, saturated soil near the surface warms while the soil below remains frozen. Water at the surface (at say 1 or 2°C) is slightly denser than the thawing water at 0°C below (water is densest at 4°C). Thus, the denser water sinks and forces the less dense water upwards. Therefore, a small convection cell can form (think of the ocean or atmospheric circulation cells described in Chapter 3). It is thought that the edges of the convection cell are associated with the surface sorted features, as soil particles may move with the water. However, the exact processes are not yet clear.

At a larger scale, **pingos** can form. These are ice-cored mounds up to 60 metres high and 500 metres in length (e.g. Figure 5.13c). The mounds contain some segregated ice and a lens of massive ice. The top of the mound often becomes cracked as the ice core within the pingo grows larger, forcing the ground surface upwards. Hydrostatic pingos are caused by the doming of frozen ground as a result of the freezing of water and the growth of permafrost beneath a former lake or other water body. The features are usually isolated landforms predominantly in areas of low relief. Pingos that form over drained lakes are usually circular in shape, whereas pingos over old river channels may be linear in form. Hydraulic pingos form most commonly in discontinuous permafrost regions at the foot of slopes and are usually circular or elliptical in shape. They result from the inflow and freezing of groundwater seeping from upslope.

If segregated ice melts at any point, then there can be a large volume of excess water and subsidence. This can lead to a terrain that is almost impassable, especially in summer, with lots of depressions, many of which are filled with water. This terrain, consisting of small, irregularly shaped thaw lakes and depressions, is called **thermokarst**.

Periglacial mass movement on slopes can result in landforms such as **protalus ramparts**, which are linear mounds of coarse sediment that form a small distance from the base of a slope. When a rock fall occurs, boulders may slide across snow at the foot of the slope, coming to rest just beyond the edge of the snow. **Ploughing boulders** move slowly downslope, leaving a depression on the hillslope indicating its path and forming a small bulge of sediment downslope of the boulder. The movement of the boulders, typically a few millimetres per year, is thought to occur because of the different thermal conditions beneath the boulder compared with its surroundings. Larger forms of ploughing boulders occur when a whole mass of rock and sediment moves downslope, a little like a glacier. These tongue shaped **rock glaciers** have a steep front and usually descend from cirques that they have created through downslope movement of the angular debris. Ice within the pore spaces assists the flow processes.

While there are no perfectly symmetrical valleys, periglacial valleys can often show a distinct asymmetry. Areas which are no longer periglacial often have asymmetric valleys that are relicts from former periglacial times. The asymmetry can be caused by south-facing slopes having longer exposure to the Sun's energy in the northern hemisphere, which promotes prolonged thawing; more frequent freeze-thaw, as there are more days and nights with freeze-thaw conditions; more melting; and more rapid mass movements. Therefore, south-facing slopes in the northern hemisphere and north-facing slopes in the southern hemisphere will experience a quicker reduction in slope angle while the deposited material at the bottom pushes the stream towards the opposite facing slope, undercutting it and keeping it steep.

SUMMARY

- The pathways that water takes through and over soils and rocks influence the response of river flow to precipitation events.
- Overland flow production by infiltration and saturation-excess mechanisms generally results in shorter lag times and higher discharge peaks in the river than in deep throughflow- and groundwater-dominated systems. However, the exact outcome can depend on soil and bedrock type, topography, vegetation cover and climatic conditions.
- Flooding and drought susceptibility have been heavily modified by human action causing changes in water flowpaths, water storage and therefore in river flow.
- River channels are dynamic in that they move position around the landscape through time, change their shape, and move sediment, water and dissolved materials.
- Human modification of river channels has led to geomorphological and ecological problems. River restoration now attempts to reverse these effects while at the same time protecting infrastructure assets.
- Natural water pathways, climate, vegetation cover, soil type, topography and management of the landscape influence how

the chemistry of precipitation is modified as it moves through the landscape into rivers, lakes and deep groundwater.

- Human action has increased the total amount of solutes transported by rivers by 12 per cent, and pollution of surface and groundwaters is widespread from point and diffuse sources, including agriculture.
- Coastal areas are dominated by wave and tidal processes that drive weathering and sediment movement.
- Coastal landforms can be characterised by wave-dominated features such as beaches, tide-dominated features such as estuaries and river-dominated features such as deltas.
- Coastal management must incorporate understanding of weathering and sediment transport processes because it is a tightly balanced system. Stopping natural sediment movements in one location on the coast may cause additional erosion and major coastal problems a little further along the coastline.
- Ice sheets and glaciers erode dramatic new landforms, such as U-shaped valleys, horns and arêtes. They also create deposits that form other more subdued landforms, such as moraines, stoss-and-lee forms and drumlin fields. These features can be seen in areas that are no longer glaciated and provide evidence of former colder climates.
- Permafrost is frozen ground, which covers 25 per cent of the Earth's land surface.
- An active layer of melt near the surface operates in the summer in permafrost areas, meaning that the ground periodically subsides.
- Infrastructure must be carefully designed in permafrost areas to cope with seasonal melt of the upper active soil layer and to prevent additional melt and subsidence of the wider permafrost caused by heating of the ground surface by infrastructure.
- Climate change is causing loss of permafrost so that infrastructure has to be designed or retrofitted to avoid significant slumping.
- Freezing water within sediment can form into large blocks of ice. Some of these can cause the surface to rise tens of metres, forming pingos.
- Frost action produces cracking, and heave and thrust processes within periglacial areas produce landforms such as polygons, stone circles, stripes, asymmetric valleys and thermokarst.

FURTHER READING

Benn, D. and Evans, D.J.A (2010) *Glaciers and glaciation* (2nd edition). London: Hodder Education.

This is a very popular book with students who want to learn more about glacial processes and features.

Boyd, C. (2020) *Water quality: an introduction* (3rd edition). New York: Springer.

An overview that gives more detail on the physical and chemical processes responsible for water quality.

French, H.M. (2017) *The periglacial environment* (4th edition). Chichester: Wiley-Blackwell.

A classic textbook, clearly written and detailed with diagrams and examples on periglacial features and permafrost.

Holden, J. (ed) (2017) *An introduction to physical geography and the environment* (4th edition). Harlow: Pearson Education.

Focus on six, well-illustrated textbook chapters by specialist experts covering more detail on 'Catchment hydrology' (pp. 465–492), 'Fluvial geomorphology and river management' (pp. 493–524), 'Solutes and water quality' (pp. 525–556), 'Coasts' (pp. 584–624), 'Glaciers and ice sheets' (pp. 625–655), 'Permafrost and periglaciation' (pp. 656–674).

Holden, J. (ed) (2020) *Water resources: an integrated approach* (2nd edition). Abingdon: Routledge.

This textbook has lots of useful, well-illustrated chapters by specialists on the changing water cycle, groundwater, surface water and water quality.

IPCC (2019) IPCC Special Report on the ocean and cryosphere in a changing climate. Pörtner, H.-O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B. and Weyer, N.M. (eds.). IPCC: Geneva.

An IPCC report covering changes to glaciers and ice sheets.

Masselink, G., Hughes, M.G. and Knight, J. (2011) *Introduction to coastal processes and geomorphology* (2nd edition). London: Hodder Education.

This text provides excellent coverage of coastal topics.

Price, M. (2002) *Introducing groundwater* (2nd edition). Cheltenham: Nelson Thornes.

This book, even after 20 years, still provides very clear coverage of practical techniques for assessing groundwater and understanding groundwater movement and water quality.

Ross, M.R.V., Topp, S.N., Appling, A.P., Yang, X., Kuhn, C., Butman, D., Simard, M., Pavelsky, T.M. (2019) AquaSat: a data set to enable remote sensing of water quality for inland waters. *Water Resources Research, 55*: 10012–10025, https://doi.org/10.1029/2019WR024883

A paper providing further details related to Box 5.3.

Shaw, E.M., Beven, K.J., Chappell, N.A. and Lamb, R. (2010) *Hydrology in practice* (4th edition). Abingdon: Taylor and Francis.

A good book aiming to provide a more detailed understanding of hydrological processes, measurement and modelling.

6

BIOGEOGRAPHY

The study of the distribution and patterns of life on Earth and of the underlying processes that result in these patterns is known as biogeography. The biosphere is the biological part of the Earth, which incorporates the Earth's surface and a shallow layer below it, the oceans and the lower atmosphere. Within the biosphere there exist many ecosystems. Ecosystems consist of the biological communities and the physical environment that sustains them where energy and nutrient cycles link the organic and mineral components of the biosphere.

THE BIOSPHERE

The biosphere is characterised by large- and small-scale energy flows and cycles of nutrients. The Earth's biosphere is not the same throughout but does have patterns of distinctive regions at all scales. Variations within the biosphere may result from factors including climate, geology, soil type, biotic processes and human action, as briefly outlined below.

KEY BIOSPHERE VARIABLES

LIGHT

Photosynthesis by green plants captures carbon from the atmosphere and combines it with water to produce complex carbohydrates and release oxygen. The carbohydrates are the building blocks of all life. The energy for photosynthesis comes from the Sun. Around one sixth of the light energy absorbed by a green

DOI: 10.4324/9781003124405-6

plant is used for photosynthesis while the rest is converted into chemical or potential food energy of the plant tissues. This energy can be used by other organisms consuming the plant tissue. The energy is released as heat through respiration in plants and animals, which consume oxygen and release carbon dioxide. Green plants therefore need light, and the more light they have the more growth can be expected. Thus, faster growth would be expected in the tropics and slower growth at the poles. The majority of plant species (C_3 plants) are found to fix carbon dioxide into '3-carbon' compounds known as triose phosphates. However, some other species (C_4 plants) make a '4-carbon' compound instead, known as oxaloacetic acid. C_4 plants have only evolved relatively recently (in the last 30 million years) within the context of life on Earth, with many C_4 plants being grasses, sedges and a few herbs and shrubs. C_4 plants have an advantage over C_3 plants in that they can utilise high levels of solar radiation effectively, use water more efficiently and are more drought tolerant. They may therefore be favoured by climate change over the next few centuries, and are among the fast-growing crops of the world, such as maize, sorghum, millet and sugar-cane.

It has only recently been discovered that light is not a prerequisite for life in the biosphere. Instead of photosynthesis, some deep-ocean ecosystems have developed **chemosynthesis**. At mid-ocean ridges (see Chapter 2), over 2 kilometres below the ocean surface where light does not penetrate, there is life. This life is sustained by energy not from sunlight, but from hot vents in the ocean floor. These hot vents emit water and many dissolved chemicals and black particles. Surrounding the vents there are large communities of animals, including tube worms several metres long and blind shrimps. Unlike photosynthesising green plants at the top of the ocean and on land, bacteria around the vents gain their energy from the chemicals released by the vents, such as hydrogen sulphide or methane. These bacteria are then grazed upon by other creatures, creating a **food chain**. Some of the bacteria even live in the shells of other creatures. The deep-ocean creatures have to avoid being boiled by temperatures from the vents, which can be over 300°C. The discovery of these ecosystems has led to a new search for life in the lightless deep lakes beneath the Antarctic ice sheet (see Chapter 5).

TEMPERATURE

The optimum conditions for growth and photosynthesis for most (but not all) plants is between 10°C and 30°C. Seasonal patterns of temperature are important as the growing season for most plants creates a baseline of food provision for other organisms. The growing season is particularly important for **herbivores** (animals that just eat plants), which must adapt to the changing availability of food resources through the seasons. They often do this by becoming dormant (e.g. hibernating) for part of the year or by migrating.

MOISTURE AVAILABILITY

Moisture availability is mainly linked to rainfall regimes. However, temperature and the ability of soils and rocks to store water that is available for biological use are also important. Soil type, geology, slope and altitude are often crucial in determining areas of increased moisture for plants and animals. All of the important plant reactions take place within water. For plants on land, water also supports their structure and without it they wilt.

OTHER CLIMATIC FACTORS

Humidity can control photosynthesis, which may not function well in very dry air. Wind can influence local temperatures. If there are strong prevailing winds in a particular location then only strong plants that are able to withstand windy conditions may grow there.

GEOLOGICAL FACTORS

The movement of tectonic plates across the surface of the Earth (see Chapter 2) has provided opportunities for species to spread or for barriers to form, such as chains of mountains or the opening of oceans. For example, there are large differences in fauna and flora between the islands of Bali and Lombok, 30 kilometres apart. Those of Bali are related closely to those on the larger islands of Java and Sumatra to the north. Those on Lombok are more like those on New Guinea and Australia to the south. Between Bali and Lombok there is an ocean trench, which has separated the plates for over 200 million years and prevented any connecting land from joining. The islands to the north of the trench were formed from the Asian continent and the islands to the south were originally part of the Australian continent. The role of plate tectonics is therefore of fundamental importance to explaining many spatial differences in plant and animal distributions. Other geological factors include soils, which are important for controlling the water available for plants and the provision of nutrients, and also the topography, which influences the receipt of the Sun's energy, the local climate and the local hydrology.

BIOTIC FACTORS

Competition for light, nutrients, water and living space, the ability to adapt and migrate, and the presence or absence of predators and prey are important components that may result in differences within the biosphere. Competition arises in situations such as at a drinking hole in a semi-arid area. Herbivores and other species may need to wait until the carnivores (meat-eaters) have left before getting a drink of water. Another type of competition occurs when there is a limited amount of food for the individuals of one species, which may lead to the exclusion of weaker individuals and the survival of the fittest, who are able to gather enough food. The **ecological niche** is the basis of most ecological patterns in the biosphere. Where there are no competitors for any of the resources required by an individual or species, the organism can occupy the ideal conditions to which it is adapted. However, because of competition, species usually have to occupy a niche that is the result of competitive interaction between several species attracted to the resources. Competition tends to be strongest between similar species, since their ecological niches are likely to overlap. The species able to survive on the lowest amount of the limiting resource will be better off.

Another biotic factor that produces geographical patterns is the isolation of groups of organisms, perhaps through plate tectonics or sea-level rise cutting off an island. The biotic component of this isolation is the lack of breeding of the species with the larger population. This means that a wider mixing of genes does not occur, and so adaptations or changes to a species can develop more quickly. Such a process can lead to the evolution of new species (**speciation**), such as with the Galapagos 'finches' studied by Charles Darwin. Thirteen species of these birds (actually not really finches but tanagers) were only found in the Galapagos islands and each had unique adaptations to their beak shapes that were suited to different food types, such as invertebrates or large seeds and nuts. This speciation allowed each of the tanager species to adopt different niches on the islands.

Organisms can also vary widely in how well they can move or spread into different areas, and this is an important determinant on the combinations of species that can be found in an area and on how the system might respond to environmental change. All organisms disperse their offspring. Some plants disperse millions of seeds, but it may be that only a few will survive. Others only release a few seeds under certain favourable conditions for growth (e.g. immediately after wildfire). These biotic processes help shape geographical patterns within the biosphere.

THE ECOSYSTEM

Ecosystems vary from huge rainforests to individual rocks. Ecosystems involve the flow of energy and nutrients within the cycle of life (e.g. Figure 6.1). This means that changes to one part of the ecosystem will affect other parts of the ecosystem. Ecosystems can be divided into several energy levels known as **trophic levels**. The lower level photosynthesisers use the Sun's energy, nutrients and water within the soil to produce organic matter. These are the primary producers. This plant matter is then eaten by herbivores who occupy the second trophic level. Some of these may be eaten by carnivores occupying the third trophic level while, in turn, some of these carnivores may be eaten by other carnivores occupying the fourth trophic level. During this whole process, waste is generated, which may be recycled back into the soil (or water). When organisms die, their remains will probably form the diet of decomposers (e.g. maggots or fungi) who transform the litter to humus (see Chapter 2). Decomposition releases the last of the energy as heat (decomposing compost heaps are often very warm).

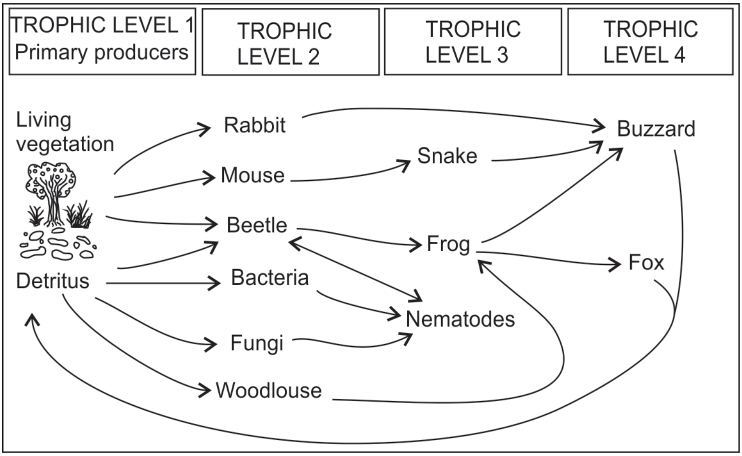

Figure 6.1 Nutrient flows in a very simplified ecosystem.

The decomposition process is important in the cycling of nutrients that have been passed through the food chain, such as nitrogen and phosphorus.

Of course, most ecosystems are more complex than the trophic levels that have just been described, but the principles remain the same. Much of the energy gained by herbivores in consuming the primary producers is used in motion, digestion, respiration and so on. Therefore, perhaps only 10 per cent of the energy is passed on from one trophic level to the next. This means that a high trophic level consumer requires a lot of primary production to support it. For a human to eat 1 kilogram of wild salmon, 1 000 kilograms of phytoplankton would need to have been produced. In addition to energy transfers, material is transferred through an ecosystem. The trophic system means that certain toxins that might be present originally in low concentrations can become concentrated in the high trophic level consumers under certain conditions. For example, a contaminant such as mercury can build up in the sediment of the sea bed and then be taken up by mussels. Each mussel may only contain a small amount of mercury. However, as small fish eat lots of mussels the mercury becomes more concentrated within the small fish. As predators eat the small fish then the mercury can accumulate to high levels within them. This can have health consequences for humans who may eat the larger fish, such as tuna and swordfish. Two further important examples of material transfer can be seen in the nitrogen and phosphorus cycles, which are described below.

PHOSPHORUS AND NITROGEN CYCLES

Rock weathering enables phosphorus to be released in a soluble form in water solution. It then becomes available within soils or water bodies for uptake by plants before being consumed by organisms at higher trophic levels. Excretion or death means that phosphorus can be taken up by decomposers and the phosphorus then becomes part of the soil or water solution again. This cycle can continue over long periods. On a larger scale, phosphorus may be washed off the land and into sea sediments. Over time this may form a sedimentary rock, which then later reaches the surface, weathers and becomes part of the water solution in soil.

Nitrogen is essential for life. Nitrogen exists in large quantities in the atmosphere. Very small amounts of nitrogen gas react with oxygen during lightning to form nitric oxide, which eventually reaches the ground as nitrate. More importantly, soil bacteria fix nitrogen to make reactive forms of ammonium and nitrate that can be used by plants. Some of these nitrogen-fixing bacteria form close relationships with plants such as legumes (e.g. clover). When used in crop rotations, these nitrogen-fixing plants can help fertilise the soil from the atmosphere, ready for the next crop. Ammonium and nitrate are taken up by plants to make protein and other parts of the plant matter. Herbivores then eat plant material to obtain their protein, and therefore their useable nitrogen. Once passed through the trophic system, dead plants and animals decompose and some of the nitrogen is acted upon by denitrifying bacteria. Nitrogen is thereby returned to the atmosphere.

Humans have disturbed the phosphorus and nitrogen cycles by developing fertilisers and also by changing the properties of soils, accelerating soil erosion and impacting on the ability of phosphorus and nitrogen to become available to plants in soil water. Furthermore, humans have also produced more nitrous oxide from industrial activity and vehicle emissions. This has resulted in enhanced nitrogen deposition from the atmosphere in rainwater, for example. Too much phosphorus or nitrogen flowing from the landscape into water bodies can cause major changes in aquatic ecosystems, as outlined in Box 6.1.

Both the phosphorus and nitrogen cycles are inter-related with parts of the carbon cycle (the carbon cycle is described in Chapter 4). For example, the productivity of plants, soil and water organisms is strongly dependent on the availability of reactive nitrogen forms, such as ammonium, nitrate and other oxidised nitrogen forms. Recent research (e.g. by Zaehle, 2013) has suggested that, at a global level, additions of reactive nitrogen to the landscape by humans increases carbon sequestration in the biosphere but also causes increases emissions of nitrous oxides (strong greenhouse gases) from soils to the atmosphere. The lack of reactive nitrogen in degraded soils is limiting terrestrial productivity in many ecosystems, and therefore limiting the capacity of the terrestrial biosphere to sequester carbon in response to increased atmospheric carbon dioxide.

BOX 6.1 EUTROPHICATION

Nitrogen and phosphorus are key limiting nutrients in aquatic ecosystems. However, increased concentrations of nitrogen and phosphorus can occur in waterbodies when fertilisers are leached from soils into watercourses or when soil erosion accelerates. The addition of nutrients stimulates the growth of primary producers. Eutrophication occurs when water is enriched with nutrients, particularly phosphorus and nitrogen, and algae thrive and become dominant (Figure 6.2). Algal blooms can block out light below, reducing the growth of plants deeper in the water body and affecting the ability of some species to prey on others. At night, when photosynthesis has stopped, the algae can lead to deoxygenation of the waterbody as both plant respiration and decomposition of detritus consumes oxygen in the water more quickly than it can be replaced by diffusion from the atmosphere. This can lead to the death of fish and other aquatic life. Often algal blooms can be seen from the air or from space due to the blue-green colouration of the surface water layers (Figure 6.2). Around a third of rivers and lakes in Europe are thought to be eutrophic due to pollution and land-use

Figure 6.2 Algal bloom indicated by the green colour in the western part of Lake Erie, North America, as observed in this natural colour image by Landsat 8 in September 2017. The image shown represents around 50 kilometres width from west to east. Some of the vast areas of farmland surrounding the lake can be seen.

(Source: NASA, https://earthobservatory.nasa.gov/images/91038/lake-erie-abloom.)

change, and it is a major threat to freshwaters and coastal marine systems globally.

Sometimes, eutrophic systems contain species of blue-green algae that produce substances that are toxic to mammals. Often these can directly affect cattle along lakes and rivers, and humans undertaking water sports. However, drinking water can also be affected. In 2014, Lake Erie's water became too toxic to treat, with a toxin called microcystin that is produced by blue-green algae found, which can cause numbness, nausea, dizziness, vomiting and can lead to liver damage. While the algal bloom was not especially larger than in other years, the wind had blown some of the bloom towards the water supply intake points. The water supply to Toledo, Ohio, and surrounding towns, was switched off for several days, affecting half a million people. Residents were advised not to even touch the water, and boiling the water was not a good remedy as this increased its toxicity. The lake has suffered from too much input of fertilisers washed from land where it has been applied to support the region's booming soya bean and maize crop systems. The result of such land management is that there have been annual algal blooms on the lake since the 1990s. Some research has also suggested that an invasive species of zebra mussel has changed the food web in the lake, which now promotes certain algal species by removing competitors. There has continued to be strong blooms on Lake Erie in the years since the Toledo water crisis (Figure 6.2). Much work needs to be done in the region, and many other areas of the world, to avoid overloading waterbodies with phosphorus and nitrogen.

SUCCESSION

Ecosystems are dynamic and are constantly adjusting to changing environmental conditions or disturbance. **Succession** occurs when older groups of plants and animals are replaced by more complex groups. Primary succession may begin on a bare rock or disturbed site. Lichens, mosses and ferns are often part of the early colonisers. Through inputs of organic matter and continued weathering and/or shelter, these plants help alter the site conditions and thereby make them more suitable for other plants and animals to colonise. These secondary colonisers, in turn, build up the complexity of the site and alter conditions further, so changing the ecosystem over time. Succession on ponds and lakes can result in the eventual infilling of the lake by organic matter and other debris.

Disturbance of an ecosystem may be natural, such as fire, a volcanic eruption, a major storm or a landslide, or may be caused by humans, such as through the felling of trees or overgrazing of domesticated animals. Disturbance is usually associated with a rapid loss of biomass. Some (but not too much) natural disturbance can be important for maintaining high biodiversity because otherwise competition between species might exclude other species.

BIOTIC INTERACTIONS

Biotic interactions need to be understood so that consequences of management decisions about ecosystems are known. *Mutualism* benefits both organisms. For example, birds eat berries while the plants benefit from the transport of their seeds. Mycorrhizal fungi exchange nutrients in the soil with plant roots, which in return provide carbohydrates for the fungi. *Predation, parasitism* and *herbivory* provide benefits to one organism at the expense of the other. In the case of predation, but also sometimes parasitism and herbivory, death of one organism will occur so that the other may survive. Parasites reduce the ability of the host organism to function by consuming nutrients from it. *Commensalism* enables one species to benefit from an interaction while the other is largely unaffected. For example, some plant seeds are adapted to be transported by animals that brush against the plant. *Amensalism* occurs when there is a negative impact to a species from an interaction while the other species is unaffected (e.g. trampling of grass by animals).

ISLAND BIOGEOGRAPHY

The study of isolated islands has yielded useful knowledge and understanding in terms of how biogeographical processes operate. Islands have clear boundaries and their isolation and lack of many external factors simplify the system and make it easier to understand. Classic island biogeography theory examines the balances between rates of immigration of new species to an island and rates of extinction on that island. If immigration rates are high and extinction rates are low then the island should be rich in species. The species richness should be related to the size of the island and how far away it is from another land mass. If an island is a short distance from a major land mass, then there will be more opportunities for new species to arrive. Larger islands are more likely to have a wider variety of habitats and so might be able to support a larger range of species. The rate of extinction of species inhabiting a new island starts off low, since competition for resources is low. As the number of species increases then the pressure on resources also increases, and so the rate of extinction rises over time. If an island is created by severance from the main continent, then the island might start off with high species richness. However, as the island has restricted resources, extinction rates will increase at first, resulting in a decline of species richness.

Island biogeography theory has also been used to help understand best management practice for conservation. For example, there have been questions about whether it is better to conserve one large area within a landscape or conserve several smaller areas. One large area might be more species rich and work well for conservation, particularly if there are large scale migratory routes through that landscape. On the other hand, while lots of small patches might mean there is less species richness within those patches, the chance of complete loss of that ecosystem (and extinction of species) is reduced because there are lots of patches with replicated ecosystems.

THE BIOMES

Biomes are global areas containing major terrestrial vegetation communities with similarities between the dominant plants and characteristic animal communities. The location of the major global biomes is shown in Figure 6.3. Climate is the main driver of the location of biomes, with temperature and precipitation regimes being crucial. The ocean is not really divided into biomes, but there are rough horizontal layers which support typical groups of plants and animals.

COLD BIOMES

The cold biomes consist of the areas including the forested taiga and the treeless tundra. Note from Figure 6.3 that these biomes are restricted in area in the southern hemisphere due to a lack of ice-free land at high latitudes. The cold biomes are therefore often called the Boreal zone (which means the northern zone).

The taiga extends north from where the monthly average temperature of 10°C occurs for less than five months of the year up to the areas where only one month has an average temperature above 10°C. The growing season is therefore short, and soils are often thin because large areas have been eroded by former ice sheets and soil development is slow due to cold temperatures. The lack of soil

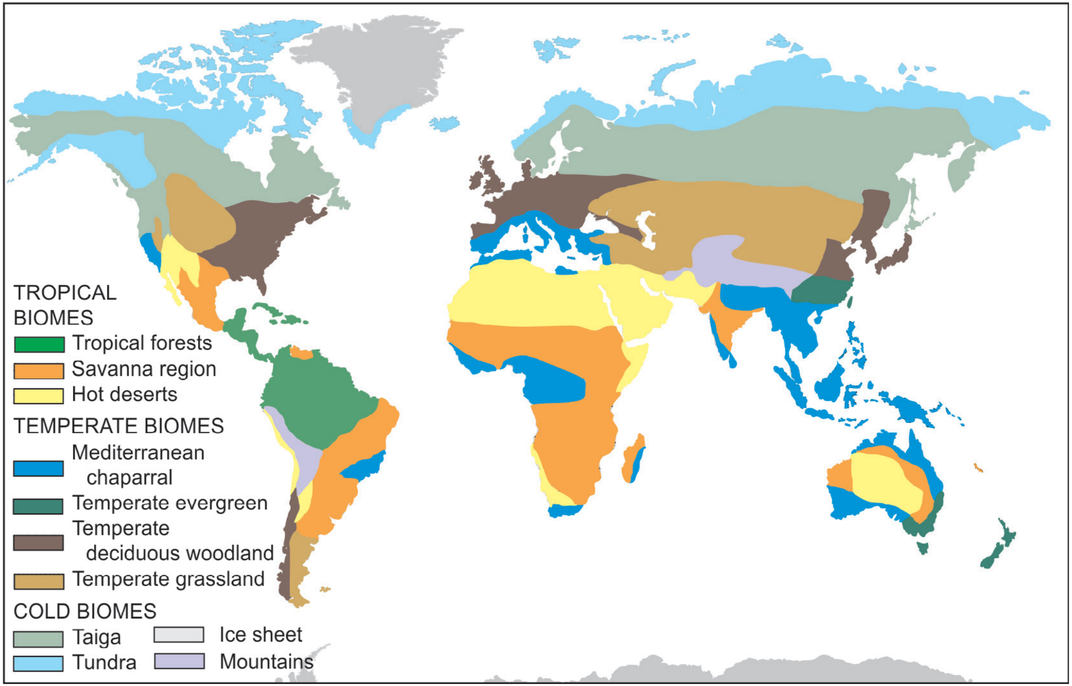

Figure 6.3 A map of the main terrestrial biomes.

animals means that decomposition is limited and there is an acidic leaf litter. In many areas, large expanses of peat have developed due to impeded drainage caused by the underlying fine glacial deposits.

The taiga forests of Europe and Asia consist mainly of coniferous trees, such as Norway Spruce and Scots Pine, in the west, but there is more deciduous larch in the east. In North America, Lodgepole Pine and Alpine Fir dominate in the west and White Spruce, Black Spruce and Balsam Fir in the east. Tree growth rates are slow, especially in the colder areas, with a height gain of perhaps 15 centimetres per decade, but evergreen species are able to photosynthesise as soon as conditions are right rather than having to wait for leaves to develop. The forest structure is simple with an evergreen canopy and then ground level tufted grasses, mosses, lichens and heathland plants such as bilberry. Tree cover is not dense everywhere and there are more open areas, especially where it is colder or where soils are very thin. Active layer processes in permafrost (see Chapter 5) can cause ground instability that can fell individual trees. Where tree cover declines, the ground is often covered with lichens. Caribou (North America) and reindeer (Europe) migrate through the taiga forest. Smaller mammals, such as weasels, tend to be camouflaged with fur that changes colour in winter.

Wildfire is an important feature of the taiga. Fire is helpful in nutrient cycling speeding up an otherwise very slow process, removing a thick, acid, litter layer (which may have been keeping more of the ground frozen for longer). Lightening strikes can initiate large fires, perhaps at 200-year intervals for a given location. These disturbances clear land to ensure there is a good species mix, but species are also adapted to fire, with many conifers, shrubs and herbs sprouting from roots, stumps and underground stems or having long seed dispersal periods (e.g. from pine cones). However, there are concerns that humans may increase fire frequency through accidental ignition or by climate change, which may occur too frequently in the future to allow for ecosystem recovery between fires.

There is a gradual transition between the taiga and the harsher tundra to the north, with trees becoming sparser in colder conditions. Tundra is the rather flat, treeless zone between the taiga and the polar ice. These regions are harsh for life: summers where temperatures rise above freezing may only last a month or two; winters may see temperatures as low as −50°C. Strong, cold, dry winds are common, with little precipitation. Permafrost with a shallow active layer inhibits plants with deep roots and reduces the number of soil animals. The soils are shallow with acid litter above a gleyed horizon and then a permanently frozen layer. There is very low productivity and species richness, with only low growing, woody, herbaceous plants and mosses and lichens. Many species are perennials, growing during the summer, dying back in the winter, and returning the following summer from their roots, thereby reducing the energy needed for seed production. The low productivity means that these areas can be susceptible to human damage as recovery rates are very slow indeed, (e.g. from vehicle tracks, mining etc.) leaving marks for hundreds of years. The low productivity also means that large areas of land are required to support any herbivores (e.g. rodents or migrating herds of reindeer and caribou) or carnivores (e.g. owls or foxes). Lemmings are important in the tundra zone as they increase the rate of nutrient cycling among the soil and plants and can even stimulate growth rates through their grazing action. They act as important prey for carnivores, but the population of lemmings changes over cycles of a few years, having knock-on effects for the rest of the ecosystem.

TEMPERATE BIOMES

The temperate biomes can be split into Mediterranean chaparral, temperate grasslands, temperate deciduous forest and southern temperate evergreen forest. The Mediterranean biome is more than just the area around the Mediterranean Sea but can be found in California, west South Africa, central Chile and southern Australia. Mediterranean chaparral climates are warm all year, with low rainfall, summer drought and high evaporation rates. The Mediterranean biome is characterised by hard, tough-leaved plants, mixed woodland and scrubland adapted to growth in conditions of limited water availability. Many plants in Mediterranean climates are adapted to frequent natural fires. Trees have thick, smooth bark and deep roots from which new growth may occur. Many types of seeds only open after exposure to fire. Animals have adapted to drought and fire, often by being able to escape quickly (e.g.

kangaroos, elk, goats and emus) or by burrowing (e.g. bobcats and rodents).

The temperate grasslands are found in extensive areas across central North America and central Euroasia. As the name suggests, the vegetation is dominated by grasses, usually perennial (the same plant surviving for year after year). There tends to be a long dry season and annual precipitation totals of usually less than 500 millimetres. The gently undulating landscapes familiar of regions such as the prairies and the steppes have often become growing areas for cereal crops. In fact, in North America there are very few areas of natural tall or short grasslands left. Large herds of herbivores, such as bison, were typical but are now reduced through human action. Leaf surface area tends to be small in temperate grasslands to reduce transpiration. Lack of a protective upper canopy cover means that animals have had to develop speed to escape from predators (e.g. antelope and deer), a large size to reduce the chances of being attacked (e.g. elk and bison) or burrowing (e.g. voles). Frequent summer fires mean that systems at or below ground level that allow survival, such as bulbs, rhizomes or tubers, are important. Scrub and trees with suckers have developed in some wetter areas of temperate grasslands, such as the South African veld or the tussock grassland found in New Zealand and Australia.

The temperate deciduous woodland biome is only found in the northern hemisphere. For similar areas in the southern hemisphere the woodland is evergreen. The differences may be related to plate tectonics, where deciduous habitats only evolved in the north after the large land mass of Gondwana split up around 180 million years ago. There is a fairly short transitional zone between the deciduous forest and the more northerly Boreal forest biome. The climate in the deciduous forest biome is moist but temperate all year, with over four months having a mean temperature above 10°C. Well mixed soils are typical, with a rich soil fauna and the soil holding plenty of nutrients. In the best areas for the forest there are four main layers to the structure of the vegetation. The upper canopy, at up to 30 metres high, has broad, rounded tree tops. There is a shrub layer below around 5 metres in height, with a third layer of grasses and a ground layer of mosses and liverworts. The vegetation cover is seasonal with dramatic changes to the appearance of the system through the year. The loss of leaves in winter reduces transpiration and frost or snow damage. Flowering of trees is usually very early in the spring, providing as much time as possible for fruits to develop. Some animals may hibernate during the more dormant period or burrow to avoid the winter cold. Deer and bears were common but humans have modified these landscapes through hunting and deforestation.

The southern evergreen forest tends to have a similar climate to the deciduous forest in the north, but generally there are two upper tree canopies and a shrub layer. As the forest is evergreen, ground level plants are less common as there are limited opportunities for good light to reach the forest floor.

TROPICAL BIOMES

The tropical biomes cover highly productive tropical rainforests, less productive savanna and low productivity hot deserts. Tropical forests tend to be consistently moist and warm, typically occurring where average annual temperatures are around 25°C with little seasonal variation, and where there is around 2 000 millimetres of regular rainfall each year. The regular water supply creates plentiful stream networks and feeds major rivers such as the Amazon and Congo. Soils in tropical forests are deep but relatively infertile, since most of the nutrients are stored within the above-ground structure of organisms, although an amazing number of soil organisms also store biomass. The cycling of nutrients is extremely rapid so plant litter does not accumulate to very deep levels, except where waterlogging promotes peatland formation. The tropical rainforests are lush with broadleaf evergreen tree cover. The tallest trees tend to be narrow with few branches or leaves below the top of the canopy. These areas produce around 40 per cent of the world's land-based primary production and contain half of the world's faunal and floral species. The lack of strong seasonality means that fruit production and growth can continue across the forest all year and there is a dense leaf canopy as the plants compete for every last bit of light. Climbers and epiphytes (plants which grow above the ground surface using other plants for support and which are not rooted in the soil) are common. Lianas (a type of climbing vine) climb rapidly and will not usually form leaves until sufficient light is available. The rainforests contain a huge diversity of animals with many adaptations for climbing, such as monkeys with strong tails, snakes and lizards. There is relatively little vegetation on the forest floor as it is so dark (only around 1 per cent of the light at the top of the canopy) but this provides room for large ground-based animals such as pigs, leopards and jaguars.

Savanna temperatures tend to be similar to those in rainforests, but the long dry season (with less than 250 millimetres rainfall per month for more than 5 months) means that the vegetation is seasonal in nature and is often adapted to be drought tolerant. The high rates of evaporation and transpiration mean that rainfall needs to be plentiful at these temperatures to achieve high productivity. A sparse tree cover allows the growth of grass and other ground flora. The structure varies across the landscape, representing local differences in water and nutrient availability. Savanna plants are adapted to withstand fire and drought, such as the baobab, which has thick bark, a short leaf season and a trunk that stores a lot of water from wet periods. Deep roots to capture water and thorns and spines to deter grazers mean that plants only need a few leaves. Fruiting of trees and other plants tends to be dominated by fire, which can occur every few years. Fruit are dropped at the end of the fire into the soil. This soil is temporarily rich in nutrients from the deposited material left after the fire. Large animals are often found in the savanna, particularly in Africa, such as wildebeest, antelope, zebra, buffalo and elephant. Nocturnal animals have adapted to reduce water loss and hide from predators (e.g. aardvark).

Hot deserts, despite their reputation, do contain plants and animals. The soils are generally poor and lack cohesion since the plant cover is extremely sparse, and organic matter inputs are therefore low. The habitat improves where water is concentrated, but most of the biomass is underground. Adaptations focus on maximising the conservation of available water; cacti store lots of water within their stems; leaves are often replaced by thorns. The creosote bush has a wide distribution of roots that put toxins into the soil to prevent other seeds, including those of the creosote, from germinating to ensure there is no local competition for water. The woodiness of many desert plants helps stop the collapse of plant material during wilting. Grasses tend to be short and tufted to protect against drought and heat. Lengthy dormant seasons are common for plants. Seeds often only germinate when the conditions are wet, perhaps many years after they have been deposited. This can mean there is a sudden bloom of life during a wet period and then offspring are not seen until the next wet event. Animals are adapted to reduce moisture loss either by being nocturnal or only producing small amounts of urine. Some animals and plants are designed to capture dew, such as the Namib beetle.

MOUNTAIN BIOMES

Topography can be an important factor influencing biomes. High mountains usually exhibit strong contrasts in the ecosystem with altitude. The zones that can be identified vary widely between mountain regions, but a typical classification may involve a hill zone where the flora and fauna is related to that in the lowlands; a montane zone, which contains species that are common to mountains but which is usually dominated by deciduous forest; a sub-alpine zone dominated by conifers with shrubs at the upper fringes; an alpine zone lacking in trees with vegetation that is short, with grasses, sedges and plants with flat, spreading growth with reduced leaves and often with colourful flowers; and a snow zone, which has sparse vegetation consisting mainly of mosses and lichens leading up to the permanently covered ice and snow zone.

AQUATIC BIOMES

The marine biome is the largest on Earth. The intertidal areas have organisms adapted to high wave energy and frequent changes as the tide varies. The strong gradient in conditions on the shore means there is a strong transition in organisms from the subtidal zone to the splash zone at the top of the shore. The species will also vary depending on whether the conditions are sandy, muddy or rocky. Where wave action is strong there will be a lack of attached plants, and animal adaptations include strong suction to enable rock attachment. The pelagic zone of the ocean occurs down to 4 000 metres depth. There is light for photosynthesis until around 200 metres depth, and here algae and bacteria provide food for fish and zooplankton. At depths greater than 200 metres, animals rely more on plankton floating downwards. Around 90 per cent of animals in the ocean from 200 to 1 000 metres depth display bioluminescence, either to attract or find prey, use as a defence or use in mating communications. The benthic zone is the area around the seafloor. It can have a large accumulation of debris and a wide variety of algae, bacteria, fungi, sponges and worms. The coldest and darkest part of the ocean, deeper than 4 000 metres, is known as the abyssal zone. There is very high pressure here, low nutrients and no plant life. However, chemosynthetic bacteria around hydrothermal vents on the ocean floor can maintain ecosystems in the zone including fish and invertebrates.

As with the oceans, freshwater ecosystems are often considered in terms of their depth zonation. While water in rivers generally mixes well, mixing in lakes is dependent on the season and their depth, and some lakes may have water that only mixes very rarely. Stratification occurs when warm water at the top of the lake traps cooler water below, and so the water does not mix well. Lakes can also stratify in winter if ice forms on the surface, resulting in warming with depth. Lake mixing can influence how nutrients are transferred as part of the aquatic ecosystem. The top of the lake (littoral zone) and shoreline can support both floating and rooted plants as light levels are good. These plants can feed zooplankton. In the deeper profundal zone, fish and invertebrates are still found, even though there is not enough light here for photosynthesis to support plants. The benthic zone can be rich in life supplied by detritus drifting to the floor from above. However, often the deeper parts of lakes can lack oxygen, which is consumed by the decomposition of detritus, and this will restrict the functioning of the ecosystem at depth. As with islands, some lakes are isolated and can contain unique species that have evolved over time in the lake ecosystem.

While river water is often well mixed, there can be ‘mesohabitats’ suited to different species. For example, pools and riffles provide quite different habitats with different flow, light and oxygen conditions, with deeper pools providing slow water velocities, finer bed sediments and lower oxygen content. Within these mesohabitats there might be smaller-scale microhabitats related to differences in flow velocity within sheltered areas, perhaps allowing the build-up of detritus, which, in turn, may influence the type and abundance of aquatic invertebrates. The conditions around the river and location along the river course may influence the aquatic ecosystem very strongly. Areas heavily shaded by trees may have reduced aquatic primary production. Hence leaf litter may provide a crucial food source for aquatic food webs. Microbes and fungi may help break down the litter while the abrasive action of flowing water and sediment can further break down the litter to smaller pieces suitable for consumption by different types of invertebrates. Shedders such as shrimps, stoneflies and caddis fly larvae can consume the organic matter derived from leaf litter. The broken down leaf litter, producing very fine particulate organic matter, can readily wash downstream, providing a food source elsewhere in the river network. In the mid reaches of rivers, fine particulate organic matter is vital for many organisms, including those that collect and filter from the water column and those that gather organic matter from the river bed. As rivers become wider, there is less shading from bankside vegetation and so primary production can increase, with more creatures that graze plants or scrape algae from rocks. However, deep pools may have less light penetration to the bed and so fine particulate organic matter may dominate in those locations, with specialist invertebrates adapted to those food sources. Predators such as fish and amphibians consume the smaller invertebrates along the river course.

Typically, there are also strong interactions between aquatic and terrestrial ecosystems. For example, terrestrial invertebrates that blow or fall into the river can be food sources for aquatic creatures, while birds, bears and bats, for example, may consume aquatic animals. Aquatic creatures may also emerge from their larval form and then become flies, which may be consumed by terrestrial creatures.

HUMAN IMPACTS

Humans have modified the biosphere by enhancing erosion and environmental pollution, over-exploiting species, causing deforestation and extinctions, accidentally (rats on ships colonising islands being visited) or deliberately (introduction of new crops to an area) distributing species, and influencing the evolution of species through the domestication of crops and animals. Furthermore, if the system is highly interconnected then the removal of one key species could have major impacts. Species that are highly connected with the rest of the food web are those whose elimination is likely to have the worst effects on the ecosystem. These are known as **keystone species** (see also Box 6.2 for impacts of reintroducing a lost keystone species). Prairie dogs are a keystone species in temperate grasslands, acting as grazers, predators, prey and providing habitats for other species through burrowing.

BOX 6.2 WOLVES AS REINTRODUCED KEYSTONE SPECIES IN YELLOWSTONE NATIONAL PARK

The grey wolf was largely eliminated by humans from Yellowstone by the 1930s. As a result, the elk population moved around less and only the severity of winters determined how many elk carcasses would be available for other species that scavenge. This meant that such food sources became less reliable. The less mobile elk also overgrazed willow and aspen trees. The reintroduction of wolves in the mid 1990s has led to elk being more mobile and on their guard, so avoiding overgrazing. Their population is kept under control and wolf kills also provide more regular food sources for scavengers and other predators such as bears. The wolves' aggression towards coyotes has benefitted rodents and birds of prey. The willow and aspen have recovered so that there are more song birds and there is less riverbank erosion. Beaver colonies have re-established where willow and aspen have been allowed to recover, benefitting aquatic biodiversity. Research is ongoing into the wider ecosystem impacts of wolf reintroduction, but the wide range of impacts on birds, trees, fish and mammals shows the importance of keystone species.

BIODIVERSITY

Biodiversity has various measures, but essentially, it is a term that describes the number and variety of species within an ecosystem. Global areas of high biodiversity usually result from a lack of major disturbance and lack of isolation. Regional and local patterns may result from short-term disturbances (fires which maintain overall ecosystem diversity) or habitat diversity. There is concern that human activity, including accelerated climate change, is causing a decline in the number of species. While 1.5 million species are currently known, there may be twice this number yet to be discovered. Since the tropical forests hold around half of the world's species, their deforestation is of major concern. Over the past 400 years, around 500 plants and 600 animals have become extinct that we know of, mainly due to human action. Natural extinction rates are normally around one mammal per 400 years and so the current rate of 89 mammals in the last 400 years far exceeds this. There may be many undiscovered species that are being driven to extinction at the moment. That is a cause for concern not only because it suggests we are damaging the ecosystems around us but also because many plants have important pharmaceutical benefits for us and if they are 'lost before they are found' then their medical uses will never be discovered. The elimination of species by direct slaughter and over-killing by humans for food, highly priced animal products such as ivory and whale oil, or to remove pests has led to major modifications of the animal kingdom. Slaughter by humans can be on an enormous scale. For example, of the 60 million bison on the Great Plains in 1700, only 21 individuals remained by 1913.

Half of the Earth's forest cover has been removed by humans. Deforestation commenced several thousand years ago accelerated by the development of the axe. The emergence of domestic livestock encouraged the clearing of land for agriculture, fuel needs grew and protection from enemies was realised through the removal of hiding places in forests. Deforestation has progressed across different zones of the Earth and its current focus is now in the tropical forests. Large amounts of carbon are stored in the biomass of the tropical forests. There is also evidence that tropical deforestation is advancing to higher altitudes than before, and some mountainous tropical regions have a high carbon density in the biomass.

It has been suggested that there are hotspots on the land surface where biodiversity is particularly high. Work by Norman Myers and colleagues (2000) has suggested that 44 per cent of all species of vascular plants are found in 25 hotspots comprising only 1.4 per cent of the land surface. Surprising areas are included as hotspots; the natural vegetation of the tropical Andes seems to be the most diverse hotspot. Even though its vegetation has been reduced to 25 per cent of its original extent by human action, it still contains 6.7 per cent of all plant species in the world and 5.7 per cent of all vertebrate animals. Myers and colleagues urge that these hotspot areas be singled out for the attention of conservationists in an attempt to protect them. It is also notable that climate change projections suggest that all of these hotspots will undergo considerable warming by 2100.

INVASIVE SPECIES

Humans are agents of the dispersal and distribution of species. The growth of long-distance transport has effectively bridged former barriers to movement for species. The deliberate introduction of wild animals from one region to another has often had unintended consequences on the ecosystem. Invasive non-native species alter ecosystem processes by changing the interactions between species. Rabbits were introduced to Australia in 1787 and 1791 and, lacking predators, were able to grow their populations at enormous rates, damaging the native vegetation cover and becoming pests. Foxes were then introduced as the natural predator to the rabbit, but instead they preferred to prey on native marsupials and birds, devastating their populations. The grey squirrel that was introduced from North America to Europe is currently causing the displacement of the native red squirrel in Britain. Many species of plants and animals (e.g. rats) have been accidentally transported in vehicles, ships and planes to become invasive species, sometimes spreading disease or driving other species to extinction. Deliberate and accidental introductions of species, which act as weed species without their natural controls of predators, alter local balances. Being careful to restrict the number of unwanted non-native introductions is therefore an important activity at airports, ports and other borders as part of biosecurity. Ships, aircraft and mail are searched and sanitised as a matter of routine by many countries. Another routine procedure requires ships to empty water-ballast tanks mid-voyage. Checking, cleaning and drying canoes or other objects that we put into rivers or lakes before transferring them to others may also help reduce the transfer of invasive species between water bodies.

AGRICULTURE

Agriculture now dominates the landscape across many areas of the world. The area dedicated to agriculture is expected to enlarge by as much as 50 per cent as the world's population grows. The agricultural process will also consume more water, which may restrict water availability for other ecosystems. Agricultural systems are ecosystems with managed inputs and outputs of energy and nutrients and with controlled species diversity. Energy flows occur along simple routes. A higher proportion of the Sun's energy is made available to crop plants, which is then passed on more directly to humans or indirectly through livestock in cropped systems than natural systems. This also means that a higher proportion of the primary production is exported from the system as the harvested product, meaning there is less organic matter and nutrients in the soil. Nutrients and organic matter are therefore added to the system via fertiliser application or by having crop rotations with fallow years. Agricultural systems tend to have low biodiversity and are fairly simple. Selective herbicides, reinforced by the use of selective fertilising or manual or mechanised weeding, reduces diversity. Further impacts of agriculture include increases in soil erosion rates (see Chapter 2) and leaching of pesticides and fertilisers into watercourses, altering aquatic ecosystems.

Humans have encouraged species evolution through agricultural processes. Normally evolution is slow. Variant individuals of a species only give rise to a new strain under conditions that favour their survival and allow them to maintain their variance by avoiding cross-fertilisation with the 'normal' members of the same species. This is naturally quite rare. Also, mixing two species to get a hybrid is naturally quite rare. This is because cross-fertilisation of the parent species dampens any variance. However, humans can create habitats that give variants a competitive advantage and, indeed, humans deliberately select plants and animals for domestication. The selection, planting and propagation of favoured plants, their variants and hybrids to suit human needs means that these plants grow well in the conditions provided for them, but many of them would not survive in the face of competition in the wild (e.g. they have lost their thorns, hairiness, toughness and so on, which is good for human consumption but bad when it comes to surviving in the wild). In fact, most crop plants currently lack the ability to reproduce or maintain themselves independently. Many food crops are entirely dependent on humans for their propagation. The banana is a sterile hybrid that produces attractive fruit but is unable to develop the seeds necessary for its own propagation. It should also be noted that around 85 per cent of food supplied to humans today is derived from less than 20 plant species, and there are concerns that these species may no longer be tough enough to withstand a new disease (imagine a COVID-19 of the plant world), and so we should be diversifying our food sources. **Genetic modification** (often called GM) is another step along the route of evolutionary change that humans have fostered in their search for more productive food crops. Instead of cross breeding or selective breeding of crops, genetic modification speeds up the process of modification by applying the changes that humans are seeking into their DNA by inserting or deleting genes. This provides a precise result rather than the less precise cross-breeding approach. Genetically modified food seems to be of concern to members of the public in some countries, although the reasons for this are not entirely clear given the long history of domestication and selective breeding.

There are relatively few species of domesticated livestock, and these tend to be herd animals which can survive on low nutrient vegetation and therefore use land that may be less suitable for crops. Pigs and cattle have been living in proximity to humans for about 10 000 years, and around 2 000 years ago, pigs, cattle, sheep, goats and buffalo had all been domesticated and had evolved into breeding groups distinct from their wild ancestors. Selective breeding has brought about changes in the physical and physiological features of these creatures. More recently, genetically modified pigs have been developed. Selective breeding really can result in fast changes to an animal species. For instance, the variety of domestic dogs we are familiar with, such as poodles, Labradors, terriers, huskies, bulldogs, Alsatians and so on all originate from one single species – the wolf – and have come about through the selective breeding of the animals, by humans, over just the last 12 000 years.

Food consumption tends to be more protein rich in developed countries with a high proportion of meat and dairy products.

Trends show that as developing countries progress, these demands also increase. This is problematic because it takes more energy, land and water to produce these foods. In fact, half of the agricultural land of the USA and Canada is currently used to provide plants for animal consumption. Therefore, not only will population increase drive changes in arable production, but the wealth of nations will change the types of production as food preferences change too. Many areas are already stretched for water resources, but an increased demand for meat and dairy products will increase water use dramatically, with impacts on other parts of the natural system.

URBAN ECOSYSTEMS

Half of the world's population lives in urban locations. These urban locations, with specific domestic animals, pests and careful or careless planting, have resulted in urban ecosystems. These ecosystems include highly managed parks and gardens with deliberate introductions of species as well as abandoned land with habitats containing native and alien species. This means that habitat diversity in urban areas can be high. The various structures produce large and small habitats that are quite different, and changes in buildings and land use means there is often change in habitats across different parts of the urban zone.

The warmer temperatures, resulting from the urban heat island effect (see Chapter 3), modified air flows and poorer air quality, compared to nearby rural areas, are components of urban ecosystems. Typically, urban areas produce more organic waste through sewage and foodstuffs than can be biologically decomposed and cycled back into the ecosystem. Some species have found the combination of resources in urban zones to be to their advantage. The rich nutrition provided by waste dumps attracts many birds and other species. House mice and house spiders, together with the brown rat, are all part of the urban nutrient cycle. Released pets have sometimes colonised urban areas (e.g. parakeets) that are not native. Predators at higher trophic levels have also come into urban areas to take advantage of new prey available in the urban ecosystems. In addition, agricultural change in rural areas, which has removed places of shelter such as shrubs and woodland, may have pushed some species into urban areas where the shelter is better.

CLIMATE CHANGE

Research has shown that vegetation zones on mountains have been moving upwards and trees expanding from the taiga further north. Spring flowering has been occurring earlier and growing seasons have been lengthening. These changes have been taking place, but it is not certain how the biosphere will respond to the rapid climate change forecast to take place by 2100 (see Chapter 4). There are simply too many plant species to examine how each one individually might respond given its characteristics. So, instead, a number of models have been designed that predict how systems might respond using **plant functional types** (plants that share similar traits and are similar in their association with environmental variables). These models seem to suggest that some tropical forests in southeast Asia, central America and the Amazon will become savanna, and some of the Mediterranean chaparral will also become savanna. The models also suggest that evergreen forest will replace grasslands in parts of North America and Northern Europe. Of course, the whole picture is quite complex because plants may grow better with more carbon dioxide in the atmosphere and also pollution that adds nitrogen to the atmosphere may be deposited on the ground, stimulating plant growth.

CONSERVATION

Conservation can mean different things to different people. An important species to protect for some may be a pest to others. Conservation motives can include ethical concerns, a desire to protect something because it looks nice and enriches our lives, a need to maintain genetic diversity, the need to keep systems complex so they are more robust to environmental change and economic incentives, such as safari tourism or the potential medicinal benefits of new food sources. Conservation management may be focussed at the ecosystem, habitat or species level. Strategies can involve legislation to ban the hunting of a certain creature or even banning human entry into special reserves. Ensuring there are corridors for species to travel between patches within ecosystems can be another strategy for conservation.

Effort can be put into conserving the *status quo* (though this may be challenging under climate change), reintroducing species or restoring functions and dynamic flows back into an ecosystem. There may also be special conservation strategies to try to ensure there is a large gene pool to protect against environmental change. Since many domesticated crops and animals would not survive in the wild, there is concern that a disease or other disaster may come along that destroys those domesticated species, and we would then not have enough of the more robust wild ancestors left from which to develop the food of the future. This latter reason for conservation is clearly about the survival of humans and has an economic incentive.

Since conservation can be emotive, there are moves to try to create rational measures by which effort and resources can be allocated. One method adopts the **ecosystem services** approach. Ecosystem services are those that support humans in some way. They may be supporting services, such as soil formation, photosynthesis, primary production, nutrient cycling and water cycling; provisioning services, such as food, fibre, fuel, chemicals, medicines and fresh water; regulating services, such as flood regulation, climate regulation, water purification, disease regulation, pollination and natural hazard regulation; and cultural services, such as recreation, spiritual enrichment, learning, reflection and aesthetic values. Evaluating the services that ecosystems provide to society in these ways can help focus attention on where investment and change might be required. For example, if water is scarce and therefore valuable, but water security could be enhanced by upstream ecosystem management, then this may aid management decisions. Investment in those changes may be worthwhile despite the fact that another service (e.g. the provision of wood for furniture) might be reduced. The economics would allow a weighing-up of whether the furniture or the water was more valuable, and payments might be forthcoming to change the management and perhaps even compensate the foresters or furniture makers, as this would still be cheaper than having to obtain water from a different location. Of course, the problem with putting things into economic terms is that there is still some subjectivity in certain areas.

For example, it is difficult to put an economic value on the cultural services of archaeological preservation or the spiritual significance of a landscape. Therefore, there needs to be a broader, balanced view taken as to how to evaluate the service provision of different ecosystems to enhance their conservation. Nevertheless, raising awareness of the wider services that ecosystems provide appears to be of benefit, and allows people to appreciate and value the services offered by quite distant systems and understand why those systems need protecting.

The **ecological footprint** is another tool for allowing people to understand the wider impact they have on the environment through the activities they undertake and the products they consume. The ecological footprint estimates the amount of ecological resources that are used by individuals, companies or countries. It is a measure of the amount of biologically productive area needed to both produce resources used and to absorb waste created by an activity or in the creation of a purchased product. Most developed countries appear to be using more resources than they can sustain and are therefore running at an ecological deficit. The USA, for example, has a large deficit, whereas Argentina has a surplus. It was estimated in 2021 by the Global Footprint Network that it takes the Earth 20 months to regenerate what we currently use in 12 months. The measure can be used to set goals for reducing consumption or increasing ecological productivity (for example, roof gardens).

SUMMARY

- Light, temperature, moisture availability, geological factors, humans and biotic factors, such as competition, adaptation, migration, mutualism, commensalism, predation and amensalism, all result in differences across the biosphere.
- Ecosystems are dynamic, with inputs of energy, cycles of nutrients and changes through time.
- The cold biomes of the tundra and taiga have low productivity and vegetation is slow-growing, restricted by short growing seasons. Irregular fire is a feature that helps to regenerate and increase productivity in the taiga.
- The temperate biomes of the deciduous forest, evergreen forest, Mediterranean chaparral and temperate grasslands are

dominated by seasonality, with the latter two showing more adaptations to fire.

- The main biomes in the tropics are the highly productive and diverse tropical forests, the lower productivity savanna and the low productivity hot deserts. Savanna species often show adaptation to regular fire.
- Humans have had a major impact on the biosphere by reducing biodiversity through deforestation, agriculture, over-exploitation of species and other environmental damage.
- Humans have assisted the movement of species around the world, both intentionally and unintentionally, often with adverse consequences, and have increased the rate of evolution for crops and livestock.
- Climate change is likely to drive shifts in biomes and provide major challenges for conservation.
- Properly evaluating the full range of services that ecosystems offer to humans provides a way of evaluating ecosystem management strategies and increases the impetus to ensure ecosystems are sustainably managed.

FURTHER READING

Begon, M., Howarth, R.W. and Townsend, C.R. (2014) *Essentials of ecology* (4th edition). Chichester: Wiley-Blackwell.

A very clear introduction to ecology.

Cox, C.B., Moore, P.D. and Ladle, R.J. (2016) *Biogeography: an ecological and evolutionary approach* (9th edition). Chichester: Wiley-Blackwell.

This is a widely used textbook.

Dickinson, G. and Murphy, K. (2007) *Ecosystems* (2nd edition). Abingdon: Routledge.

An excellent textbook covering ecosystem concepts, such as energy and material flows, ecosystem disturbance, succession and human impacts. There is also a very good section on biomes.

Holden, J. (ed) (2017) *An introduction to physical geography and the environment* (4th edition). Harlow: Pearson Education.

Focus on four, well-illustrated, textbook chapters by specialists covering more detail on 'The biosphere' (pp. 253–276), 'Ecosystem Processes' (pp. 277–297), 'Freshwater ecosystems' (pp. 298–322) and 'Vegetation and environmental change' (pp. 323–343).

Moss, B. (2018) *Ecology of freshwaters: Earth's bloodstream* (5th edition). Chichester: Wiley-Blackwell.

Comprehensive textbook demonstrating the integrated nature of freshwater systems.

Wilson E.O. (2010) *The diversity of life* (2nd edition, with a new preface). London: Belknap Press.

This award-winning book provides excellent bed-time reading and is also highly informative.

Zaehle, S. (2013) Terrestrial nitrogen–carbon cycle interactions at the global scale. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*: 20130125.

A research paper illustrating how modelling can help us understand ecosystem processes and cycling of nutrients.

7

SOLUTIONS IN PHYSICAL GEOGRAPHY

With ever-increasing demands for food, energy, water, minerals and rare metals from a growing and developing world population, there is a need to develop solutions for major global challenges as the world's environment changes and as climate change adds to migration and wellbeing challenges created by wars. Using technology such as remote sensing techniques (e.g. see Boxes 2.1, 5.1, 5.3 and 5.6) and sophisticated automated warning systems informed by artificial intelligence (e.g. see Boxes 2.2 and 5.2) can provide important support for managing hazards. Such systems are operational and active every day. Monitoring and modelling (e.g. Boxes 3.1 and 4.1, sections on IPCC in Chapter 4 and climate change section in Chapter 6) can support physical geographers in making predictions about the inter-related feedbacks within the Earth's system so that we can understand how environmental change may progress and how different scenarios for environmental management might play out. These existing tools, and new tools that are developed each day, inform policy makers and environmental managers in the development of suitable approaches for managing environmental change.

MANAGING ENVIRONMENTAL HAZARDS

There is a wide range of natural hazards. Some have become more hazardous to humans because of how populations have concentrated in key areas (e.g. floodplains, coastal zones), while others have potentially been made more threatening due to how humans have affected the Earth's climate (e.g. wildfire, storms, sea-level

DOI: 10.4324/9781003124405-7

rise) or landscapes (e.g. deforestation potentially exacerbating flood risk downstream). Humans have also come up with effective solutions to managing some hazards (e.g. tsunami warnings, water quality treatment).

Environmental hazards result from an event or substance that can lead to harm or disruption to humans. There are physical hazards, such as those associated with volcanic eruptions, biological hazards, including infectious diseases like COVID-19, and chemical hazards, such as air or water pollution. Hazards may also have multiplier effects (creating compound events) as they interact with one another. For example, a drought may lead to soil surface crusting and reduced infiltration capacity, and perhaps also wildfire that removes surface vegetation. A rainstorm following such a drought and wildfire may then lead to enhanced flooding due to the reduced infiltration capacity and the lack of vegetation to slow the flow of water. Additionally, landslides may occur due to the vegetation removal and surface soil erosion may increase, while pollution of water bodies may be enhanced as surface pollutants are washed from the land surface. Understanding relationships between different hazards is therefore important when assessing risks and taking action. This understanding requires researchers from different fields to work together.

Scientists have developed interdisciplinary solutions, which typically involve increasing community preparedness and building resilience into infrastructure or society. There are mitigation solutions and adaptation solutions. Mitigation and adaptation measures are usually undertaken together. The former involves reducing vulnerability to a risk by changing the circumstances, such as undertaking spatially designed natural flood management measures in the headwaters of a catchment to reduce downstream flooding, or enforcing the use of buffer zones and appropriate timing of fertiliser applications to reduce the amount of nitrate leaching into watercourses. Adaptation does not address the cause of a hazard but is focused on minimising the impacts. For example, businesses in a flood zone may have rapidly moveable shelving so that products can be moved to higher levels when there is a flood warning. Trust of businesses and communities is enhanced when the flood warnings are accurate and well communicated, and researchers therefore use both real-time models and monitoring and long-term scenario modelling in an attempt to continuously improve and refine flood predictions.

Earthquake or volcano early warning systems using remote sensing technologies that measure small changes in the Earth's surface, and assemblages of weather models that are run in tandem to produce better storm predictions are examples of solutions used to enhance community response to hazards and reduce the risks. However, it is still the case that disadvantaged communities suffer more from natural hazards, with a higher proportion living on floodplains, in poorer quality buildings less able to withstand earthquakes, and less able to move when warnings of an impending hazard are declared. Thus, researchers and policy makers need to do even more to produce workable solutions that support the most vulnerable in society. It is not sufficient to develop smartphone hazard warning apps if the most vulnerable do not own a smartphone. The people themselves need to be involved in developing the solutions and enhancing resilience. Thus, social engagement is critical for dealing with hazards and environmental change.

DEALING WITH ENVIRONMENTAL CHANGE

We know that the climate will change. Study of the Earth's landforms shows us how they have been shaped by ice sheets, glaciers and periglacial activity many times in the past in areas where there is no ice today. Sea levels have risen and fallen in tandem with the retreat and advance of the ice. Even without human intervention there will be climate change. However, the changes brought about by humans to the Earth's atmospheric composition and its climate are exceptional. Equally, the changes brought about by humans to ecosystems, soils, water pathways, river flow and channel change, water quality and coastal sediment dynamics are huge. It is hard to see how these changes will not be compounded in the near future as there are major challenges ahead. For example, we need to sort out the food and water supply for the world's growing population as it rises to 9 billion by around 2050. This needs to be done without devastating lots of ecosystems, which might deliver fundamental services and biodiversity. We need to tackle climate change and its impacts through mitigation and adaptation. These challenges require international co-operation in ways we have not witnessed before.

Co-operation is needed because the Earth has a climate system that affects all nations. Technology, resources and trade mean we live in a global world, with a global economy. Carbon and food production is often in quite different locations from where products are used/consumed. The need for co-operation was exemplified by a theory called the **tragedy of the commons**, based on ancient stories, outlined by Garrett Hardin in the journal *Science* in 1968 (Hardin, 1968). This theory involves a field that anyone can use. Each farmer would be expected to keep as many cattle as possible on the field for their own gain. However, the logic of this common resource will bring tragedy. Each farmer seeks to maximise their gain and thinks about adding one more cow to the herd. However, this action has a negative and a positive effect. The positive effect is that the farmer earns money from that additional animal. However, the negative effect is the additional overgrazing that is created by that extra animal. Because the effects of overgrazing are shared by all the farmers using the field, the negative effect for the one farmer is, therefore, minor. The problem is that the farmer may think about adding more animals, coming to the same conclusion each time. Indeed, all of the farmers will think in a similar, rational way. This leads to the tragedy. Each farmer exists within a system that encourages them to increase their herd without limit but in a field with limited resources. The field will be massively overgrazed and all of the vegetation will be removed. Therefore, none of the animals will survive and the farmers will be ruined.

This tragedy of the commons example may seem odd, as you would think that the farmers would realise that there is a limited resource in the field and, collectively, would then try to sustainably share the resource provided by the field. Indeed, this is often the case and collective or institutional management with careful rules or traditions that people are expected to follow can avoid over-exploitation. However, it is not such an odd example since it is quite analogous to allowing people or countries the freedom to pump pollutants into the atmosphere or oceans when and where they want, and in whatever quantity they want, as if the atmosphere and oceans are unlimited resources. We have finite resources and we must therefore manage those resources. Communicating the nature of the problem to all those who may be affected is essential. This is where the role of physical geography comes in. The study of physical geography can be used to inform environmental managers and politicians of potential threats to the environment, and the potential solutions to protect and enhance the environment (such as how to increase soil carbon content on farmland, which holds in more nutrients and moisture). Physical geography provides the scientific insight and understanding of the inter-connected components of the Earth's system. It allows us to understand what resources we have and how they are affected by different processes occurring on different parts of the planet.

The environment is dynamic and we know from the material covered in this book that complex feedback mechanisms operate on Earth that sometimes throw up surprising rates and directions of change. A gradual slow change, which allows lots of time to think about a management solution, may suddenly become a rapid change if the system has reached a threshold and suddenly jumps out of one stable state (see Chapter 6). This is a bit like a river that suddenly shifts its course due to a gradual build up of sediment clogging the channel, which no longer provides an optimum route for flow. For another tipping point example in the Earth's system, look at the discussion on the thermohaline circulation system in Chapter 2. These examples tell us that the physical environment and environmental change is often uncertain. However, as a solution, management action is needed early, probably well before there is a clear understanding of how things are going to change. This is known as the **precautionary principle**. There is risk involved in taking early action because it might physically be the wrong action, and it may also be costly to implement and may not have been required. However, often the costs and physical consequences of not taking early action are enormous.

Research in physical geography has told us an enormous amount about how the world works. It can inform management and policy-making by providing insights into the integrated way in which the Earth's surface processes operate and the feedbacks that often exist. The understanding we have gained from research on the Earth's systems has led to the international meetings on climate change; action on new energy technologies; the quick international action in the 1980s and 1990s on CFCs to deal with the Antarctic ozone hole; and laws to tackle over-fishing, protect soils, develop integrated land management solutions to floods, water quality, coastal management and so on. While we have been damaging the environment for quite some time, we have now taken notice, and international effort is being put into tackling the issues, including finding new geo-engineering solutions (see Chapter 4).

SUPPORTING THE SUSTAINABLE DEVELOPMENT GOALS

Just as the Earth's climate system, **geomorphological** processes and biosphere are inter-related, the world's major challenges around climate change, energy provision, poverty, malnutrition, disease and hostile feelings around differences in the standard of living between countries are inter-related. Empowerment comes from understanding how human–environment relationships are inter-related, and enables governments and people to take action thoughtfully, minimising negative feedback effects.

The need to understand processes that operate close to the Earth's surface, in order to provide information for environmental management in support of human sustainable development, should be a major driver of research. The United Nation's Sustainable Development Goals (SDGs) were adopted in 2015 as targets to be achieved by 2030. There are 17 core goals with many sub-targets under each goal. The core targets include, for example, the eradication of hunger; good health and wellbeing; gender equality; clean water and sanitation; affordable and clean energy; responsible consumption and production; climate action; and protection for life below water and life on land. The SDGs seek to drive activity to end poverty while providing environmental protection. It is thought that to achieve the SDGs it will cost $4 trillion per year, a cost probably exacerbated by the COVID-19 global pandemic. From a physical geography perspective, solutions to the eradication of hunger are being developed through, for example, technological developments in precision agriculture (Chapter 2), remote sensing of soil and crop conditions, research on net zero agriculture, soil functions and hydrological processes, ecosystem dynamics, biogeochemical cycling and rock weathering (see for example Box 4.3).

MAKING NEW DISCOVERIES

The focus of this brief concluding chapter has been about the applied nature of physical geography, the need for collaborative environmental management and the need for research to inform management action. However, I end on a note that says we also need serendipitous physical geography. I write as a research scientist (indeed as Chair of Physical Geography at the University of Leeds) who is passionate about making sure that science has an impact on society and the environment. However, we must also allow curiosity-driven research into environmental processes to thrive. Often, many amazing discoveries are made that have huge impacts on society when there is a thriving research community able to follow new avenues and find out how the world works without there necessarily being a particular societal impact in mind. The inventors of the laser were just investigating some of Einstein's theories and did not foresee its practical use for eye surgery, DVD players or geomorphological mapping. Microwave ovens and penicillin were outcomes of research on something completely different from their use today. Some Japanese researchers were recently studying the structure of some enzymes, but when they modified one enzyme to allow detailed investigation of its structural properties they found they had accidentally modified it so that it became better at breaking down some forms of plastic. Such an achievement could be an important future solution to dealing with plastic waste in the environment. Researchers at the Mote marine laboratory in Florida established some coral nurseries as part of research into coral growth. As corals grow very slowly, it seems very challenging to think that we can actively restore damaged systems. However, Mote researchers accidentally found that small fragments of coral can quickly grow towards one another and join together, creating large areas of growth each month at a much faster rate than individual coral larvae can spread (Forsman et al., 2015). They are now using this discovery as a restoration solution by breaking corals into small fragments and then letting them grow together into larger colonies. These colonies can then be planted into the ocean to help restore damaged systems. A final example of an accidental discovery is DNA fingerprinting. This allows individuals to be identified and the family tree to be traced. From an environmental perspective, such a discovery allows researchers to examine the populations of animals to determine the success of conservation measures, including monitoring family history, role of individuals, breeding patterns or just the presence/absence of endangered or invasive species (for example by tracing DNA in animal droppings or extracting it from soil and water samples), and then to intervene based on the findings. The technique also has many other applications, including the study of ancient environments.

I hope, therefore, that the above examples, and this book as a whole, can encourage you to further study physical geography not just because you might *need* to know how components of the Earth's system work for your occupation or to pass an exam, but because you *want* to know how the Earth's system works. You might just make a serendipitous discovery that benefits the environment and/or provides a solution to environmental change and hazards that humans face. In doing so, you may benefit many human beings in the long term and support the ongoing delivery of the SDGs.

SUMMARY

- Physical, biological and chemical environmental hazards can occur independently or in combinations, sometimes resulting in compound events.
- Mitigation and adaptation solutions to hazards and environmental change are usually undertaken together.
- Both interdisciplinary and stakeholder engagement approaches to delivering solutions on environmental hazards and environmental change are often required.
- The Sustainable Development Goals are driving national and international governmental and non-governmental organisations to work towards common outcomes to eradicate poverty whilst protecting the environment.
- Societal and environmental impact-focused research has yielded many environmental solutions, but serendipitous, curiosity-driven research should also be allowed to flourish as can also yield major solutions to grand challenges in the long term.

FURTHER READING

Forsman, Z.H., Page, C.A., Toonen, R.J. and Vaughan, D. (2015) Growing coral larger and faster: micro-colony-fusion as a strategy for accelerating coral cover. *PeerJ, 3*: e1313, https://doi.org/10.7717/peerj.1313.

An article describing the proof and benefits of an accidental discovery about how to accelerate coral growth as described in the final section of this chapter.

Holden, J. and McDonald A.T. (2017) Dealing with hazards and environmental change. In Holden, J. (ed) *An introduction to physical geography and the environment* (4th edition) (pp. 719–735). Harlow: Pearson Education.

A chapter that outlines key principles and approaches to environmental management.

Pine, J.C. (2014) *Hazards analysis: reducing the impact of disasters.* Boca Raton Florida: CRC Press.

This practical book illustrates hazard identification, risk analysis, management and resilience.

REFERENCES

Bhaskar, A.S., Hopkins, K.G., Smith, B.K., Stephens, T.A. and Miller, A.J. (2020) Hydrologic signals and surprises in U.S. streamflow records during urbanization. *Water Resources Research*, 56: e2019WR027039.

Cao, C., Lee, X., Liu, S., Schultz, N., Xiao, W., Zhang, M., Zhao, L. (2016) Urban heat islands in China enhanced by haze pollution. *Nature Communications*, 7: 12509.

Evans, C.D.et al. (2021) Overriding water table control on managed peatland greenhouse gas emissions. *Nature*, 593, 548–552.

Fawzy, S., Osman, A.I., Doran, J., Rooney, D.W. (2020) Strategies for mitigation of climate change: a review. *Environmental Chemistry Letters*, 18: 2069–2094.

Forsman, Z.H., Page, C.A., Toonen, R.J. and Vaughan, D. (2015) Growing coral larger and faster: micro-colony-fusion as a strategy for accelerating coral cover. *PeerJ*, 3: e1313, https://doi.org/10.7717/peerj.1313.

Hardin, G. (1968) The Tragedy of the Commons. *Science*, 162: 1243–1248.

Hubau, W.*et al.* (2020) Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature*, 579: 80–87.

IPCC. (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

IPCC (2019) IPCC Special Report on the ocean and cryosphere in a changing climate. Pörtner, H.-O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B. and Weyer, N.M. (eds.). IPCC: Geneva.

Kelland, M.E.et al. (2020) Increased yield and CO2 sequestration potential with the C4 cereal Sorghum bicolor cultivated in basaltic rock dust-amended agricultural soil. *Global Change Biology*, 26: 3658–3676.

King, M.D. et al. (2020) Dynamic ice loss from the Greenland Ice Sheet driven by sustained glacier retreat. *Communications Earth and Environment*, 1: 1.

Lewis, S.L. and Maslin, M.A. (2015) Defining the Anthropocene. *Nature*, 519: 171–180.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A. and Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, 403: 853–858.

Ritchie, H. (2020) Our World in Data, https://ourworldindata.org/ghg-emissions-by-sector#note-2. Creative Commons CC BY 4.0: https://creativecommons.org/licenses/by/4.0)

Ross, M.R.V., Topp, S.N., Appling, A.P., Yang, X., Kuhn, C., Butman, D., Simard, M., Pavelsky, T.M. (2019) AquaSat: a data set to enable remote sensing of water quality for inland waters. *Water Resources Research*, 55: 10012–10025, https://doi.org/10.1029/2019WR024883

Shepherd, A.et al. (2020) Mass balance of the Greenland Ice Sheet from 1992 to 2018. *Nature*, 579: 233–239.

Steffen, W., Richardson, R., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., de Vries, W., de Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M., Persson, L.M., Ramanathan, V., Reyers, B., Sorlin, S. (2015) Planetary boundaries: guiding human development on a changing planet. *Science*, 347: 1259855.

Voss, K.A., Famiglietti, J.S., Lo, M., de Linage, C., Rodell, M., Swenson, S. C. (2013) Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resources Research*, 49: 904–914.

Woosley, R.J. and Millero, F.J. (2020) Freshening of the western Arctic negates anthropogenic carbon uptake potential. *Limnology and Oceanography*, 65: 1834–1846.

Zaehle, S. (2013) Terrestrial nitrogen–carbon cycle interactions at the global scale. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368: 20130125.

INDEX