

2ND EDITION

Basics of Environmental Science



MICHAEL ALLABY

Also available as a printed book
see title verso for ISBN details

Basics of Environmental Science

Basics of Environmental Science is an engaging introduction to environmental study. The book offers everyone studying and interested in the environment, an essential understanding of natural environments and the way they function. It covers the entire breadth of the environmental sciences, providing concise, non-technical explanations of physical processes and systems and the effects of human activities.

In this second edition, the scientific background to major environmental issues is clearly explained. These include global warming, genetically modified foods, desertification, acid rain, deforestation, human population growth, depleting resources and nuclear power generation. There are also descriptions of the 10 major biomes.

Michael Allaby is the author or co-author of more than 60 books, most on various aspects of environmental science. In addition he has also edited or co-edited seven scientific dictionaries and edited an anthology of writing about the environment.

Basics of Environmental Science

2nd Edition

Michael Allaby



London and New York

First published 1996
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

Second edition 2000
Routledge is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2002.

© 1996, 2000 Michael Allaby

The right of Michael Allaby to be identified as the Author of this Work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

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

Library of Congress Cataloguing in Publication Data
A catalog record for this book is available from the Library of Congress

ISBN 0-415-21175-1 (hbk)
0-415-21176-X (pbk)
ISBN 0-203-13752-3 Master e-book ISBN
ISBN 0-203-17969-2 (Glassbook Format)

Contents

List of Figures	vii
List of Tables	xii
Preface to the Second Edition	xiii
How to Use This Book	xiii
1 Introduction	1
1. What is environmental science?	1
2. Environmental interactions, cycles, and systems	4
3. Ecology and environmentalism	7
4. History of environmental science	10
5. Changing attitudes to the natural world	13
Further reading	17
Notes	17
References	17
2 Earth Sciences	19
6. Formation and structure of the Earth	19
7. The formation of rocks, minerals, and geologic structures	23
8. Weathering	27
9. The evolution of landforms	30
10. Coasts, estuaries, sea levels	34
11. Energy from the Sun	37
12. Albedo and heat capacity	42
13. The greenhouse effect	44
14. The evolution, composition, and structure of the atmosphere	51
15. General circulation of the atmosphere	54
16. Oceans, gyres, currents	59
17. Weather and climate	64
18. Glacials, interglacials, and interstadials	68
19. Dating methods	73
20. Climate change	76
21. Climatic regions and floristic regions	81
Further reading	86
Notes	87
References	87
3 Physical Resources	90
22. Fresh water and the hydrologic cycle	90
23. Eutrophication and the life cycle of lakes	95
24. Salt water, brackish water, and desalination	99
25. Irrigation, waterlogging, and salinization	103
26. Soil formation, ageing, and taxonomy	107
27. Transport by water and wind	111
28. Soil, climate, and land use	115
29. Soil erosion and its control	119
30. Mining and processing of fuels	123
31. Mining and processing of minerals	130
Further reading	135
Note	135
References	135

4	Biosphere	137
32.	Biosphere, biomes, biogeography	137
33.	Major biomes	141
34.	Nutrient cycles	147
35.	Respiration and photosynthesis	151
36.	Trophic relationships	151
37.	Energy, numbers, biomass	160
38.	Ecosystems	163
39.	Succession and climax	168
40.	Arrested successions	172
41.	Colonization	176
42.	Stability, instability, and reproductive strategies	179
43.	Simplicity and diversity	183
44.	Homoeostasis, feedback, regulation	188
45.	Limits of tolerance	192
	Further reading	197
	References	197
5	Biological Resources	200
46.	Evolution	200
47.	Evolutionary strategies and game theory	206
48.	Adaptation	210
49.	Dispersal mechanisms	214
50.	Wildlife species and habitats	218
51.	Biodiversity	222
52.	Fisheries	227
53.	Forests	233
54.	Farming for food and fibre	239
55.	Human populations and demographic change	249
56.	Genetic engineering	250
	Further reading	257
	Notes	257
	References	258
6	Environmental Management	261
57.	Wildlife conservation	261
58.	Zoos, nature reserves, wilderness	265
59.	Pest control	269
60.	Restoration ecology	274
61.	World conservation strategies	237
62.	Pollution control	281
63.	Hazardous waste	287
64.	Transnational pollution	288
	Further reading	296
	References	296
	End of book summary	298
	Glossary	300
	Bibliography	307
	Index	316

Figures

2.1	Structure of the Earth	20
2.2	Plate structure of the Earth and seismically active zones	22
2.3	The mountain-forming events in Europe	25
2.4	Stages in the development of an unconformity	26
2.5	Gradation of clay and sand to laterite	29
2.6	Slope development	32
2.7	Drainage patterns	33
2.8	Deposition of sand and formation of an estuarine sand bar	35
2.9	The development of a sea cliff, wave-cut platform, and wave-built terrace	37
2.10	Average amount of solar radiation reaching the ground surface	39
2.11	Absorption, reflection, and utilization of solar energy	40
2.12	The greenhouse effect	45
2.13	Anticipated changes in concentration of three greenhouse gases	47
2.14	IPCC estimates of climate change if atmospheric CO ₂ doubles	48
2.15	Structure of the atmosphere	52
2.16	Chemical composition of the atmosphere with height	55
2.17	Seasons and the Earth's orbit	56
2.18	General circulation of the atmosphere	58
2.19	The development of cells in jet streams and high-level westerlies	58
2.20	Weather changes associated with El Niño-Southern Oscillation events	60
2.21	Ocean currents	62
2.22	Formation of cloud at a front	67
2.23	Distribution of cloud around frontal systems	67
2.24	Parts of the Earth covered by ice at some time during the past 2 million years	70
2.25	Temperature changes since the last glacial maximum	71
2.26	Orbital stretch	77
2.27	Wobble of the Earth's axis	77
2.28	Variations in axial tilt (obliquity of the ecliptic)	78
2.29	World climate types	82
2.30	Floristic regions	84
3.1	Water abstraction	91
3.2	Principal cities bordering the Rhine	93
3.3	The Rhine basin, draining land in six countries	94
3.4	The life cycle of a lake	98
3.5	Evolution of a lake into dry land, marsh, or bog	99
3.6	Multistage flash evaporation	102
3.7	Mole drainage	105
3.8	Saltwater intrusion into a freshwater aquifer	108
3.9	Soil drainage	108
3.10	Profile of a typical fertile soil	109
3.11	Flood plain development from meander system	114
3.12	Modern soil developed over flood plain alluvium and glacial till	114

3.13	Profiles of four soils, with the vegetation associated with them	116
3.14	World distribution of soil orders	118
3.15	Two types of terracing for reducing runoff	122
3.16	Effect of a windbreak in reducing wind speed	123
3.17	Types of coal mines	124
3.18	Structural oil and gas traps	126
3.19	Blast furnace and steel converter	133
4.1	Biomes and climate	139
4.2	Marine zones and continental margin	140
4.3	The nitrogen cycle	148
4.4	The carbon cycle	149
4.5	Photosynthesis	154
4.6	Simplified food web in a pond	158
4.7	Simplified heathland food web	159
4.8	Pyramid of numbers per 1000 m ² of temperate grassland	161
4.9	Flow of energy and nutrients	162
4.10	Ecosystem	165
4.11	Forest stratification	167
4.12	Succession to broad-leaved woodland	169
4.13	Succession from a lake, through bog, to forest	170
4.14	The effect of fire on species diversity	173
4.15	Effect of grazing on succession	175
4.16	Establishment of colonizers in an area of habitat	177
4.17	Island colonization as a ratio of immigration to extinction	178
4.18	Population growth and density	181
4.19	J-and S-shaped population growth curves	182
4.20	Resilience and stability	186
4.21	The edge effect	187
4.22	Speed governor of a steam engine	189
4.23	Feedback regulation of a population	190
4.24	Density-dependent feedback regulation	191
4.25	Limits of tolerance and optimum conditions	193
4.26	Plant response to temperature	195
5.1	Effects of natural selection	204
5.2	Mendelian inheritance	205
5.3	The Prisoner's Dilemma	206
5.4	Optimum foraging strategy	208
5.5	Adaptive radiation of Darwin's finches	211
5.6	Adaptation by mangroves to different levels of flooding	212
5.7	Common pattern for passive dispersal	215
5.8	Expansion of the European starling's range in North America 1915–50	218
5.9	Habitats in a pond	220
5.10	Population size needed for a 95 per cent probability of persisting 100 years	221
5.11	Species richness	225
5.12	Range and population increase	226
5.13	World fisheries catch (marine and freshwater) 1972–92	228
5.14	North Sea herring stocks 1960–90	230
5.15	Commercial fishing methods	231
5.16	Percentage of land area under forest in various countries	234

5.17	Tree cover in the British Isles about three thousand years ago	236
5.18	Traditional tree management	237
5.19	Ploughing and sowing	240
5.20	Indices of per capita food production 1990–94	243
5.21	World production of cereals during the 1990s	244
5.22	Rate of world population growth	246
5.23	World population 1850–2025 (median estimate)	248
5.24	Estimates of the rate of global population increase since 1975	249
5.25	One method of genetic engineering	252
6.1	Effects on a population of fragmentation of habitat	261
6.2	Population structure for three species within a habitat	263
6.3	Island wildlife refuges	267
6.4	Pesticide use and crop yield	270
6.5	Even-sized droplets from the teeth of an ultra-low-volume pesticide sprayer	271
6.6	A hand-held ultra-low-volume sprayer	272
6.7	Florida, showing the location of the Everglades	275
6.8	Living resources and population	278
6.9	Resource consumption by rich and poor	278
6.10	Kondratieff cycles	280
6.11	Government assistance for environmental technologies in the EU 1988–90	284
6.12	Private investment in pollution control during the 1970s and 1980s	285
6.13	Carbon dioxide emissions in 1988	286
6.14	Acid rain distribution	290
6.15	Countries bordering the Mediterranean	292
6.16	Areas included in the UNEP Regional Seas Programme	293

Tables

2.1	Albedos of various surfaces	43
2.2	Effect of the incident angle of radiation on water's albedo	43
2.3	Average composition of the troposphere and lower stratosphere	54
2.4	Geologic time-scale	74
3.1	Composition of sea water	101
3.2	Ions in sea water	101
4.1	Minerals in an oak forest as a proportion of the total	148
4.2	Items making up the diet of the blackbird <i>Turdus merula</i>	157
5.1	Number of species described and the likely total number	224
5.2	The 20 most important species in the world's fish catch	228

Preface to the Second Edition

Three years have passed since the first edition of *Basics of Environmental Science* appeared. During this time new concerns have arisen, the controversy in Britain over the safety and desirability of genetically modified foods being the most spectacular example. At the same time, our understanding of other issues has improved as more information about them has been gathered.

Revising the book for its new edition has given me the opportunity to add more information where it is now available and to outline some of the new controversies, including that over genetically modified food. At the same time I have been able to study the whole of the text and to bring it up to date where necessary.

At intervals throughout the book I have added links to sites on the World Wide Web. This has now become an invaluable educational resource and I am delighted to have been able to weave this book into its fabric.

Revised, updated, and modernized, I hope that the new edition will be of value and interest to everyone seeking to broaden their understanding of the science behind environmental issues.

Michael Allaby
Wadebridge, Cornwall
November 1999

How to Use This Book

Basics of Environmental Science will introduce you to most of the topics included under the general heading of ‘environmental science’. In this text, these topics are arranged in six chapters: Introduction; Earth Sciences; Physical Resources; Biosphere; Biological Resources; and Environmental Management. Within these chapters, each individual topic is described in a short section. There are 62 of these sections in all, numbered in sequence. All are listed on the contents pages.

You can dip into the book anywhere to read a chapter that interests you. Each is self-contained. It is not quite possible to avoid some overlap, however. This means you may find in one section a technical term that is not fully explained. In the section ‘Energy from the Sun’ (section 11), for example, you will come across a mention of the ‘greenhouse effect’, but without a detailed explanation of what that is. When you encounter a difficulty of this kind, refer to the contents pages. In this example you will find a section, number 13, devoted to the ‘greenhouse effect’, in which the phenomenon is explained fully. If there is no section specifically devoted to the term you find troublesome, look in the index. Almost certainly the term will be explained somewhere, and the index will tell you where to look. Some of the terms that you may find less familiar are defined in the glossary.

At the end of each chapter you will find a list of sections that contain explanations of terms you have just encountered.

This procedure may seem cumbersome, but it would be impractical to provide a full explanation of terms each time they occur.

1

Introduction

When you have read this chapter you will have been introduced to:

- a definition of the disciplines that comprise the environmental sciences
- cycles of elements and environmental interactions
- the difference between ecology and environmentalism
- the history of environmental science
- attitudes to the natural world and the way they change over time

1 What is environmental science?

There was a time when, as an educated person, you would have been expected to converse confidently about any intellectual or cultural topic. You would have read the latest novel, been familiar with the work of the better-known poets, have had an opinion about the current state of art, musical composition and both musical and theatrical performance. Should the subject of the conversation have changed, you would have felt equally relaxed discussing philosophical ideas. These might well have included the results of recent scientific research, for until quite recently the word ‘philosophy’ was used to describe theories derived from the investigation of natural phenomena as well as those we associate with philosophy today. The word ‘science’ is simply an anglicized version of the Latin *scientia*, which means ‘knowledge’. In German, which borrowed much less from Latin, what we call ‘science’ is known as *Wissenschaft*, literally ‘knowledge’. ‘Science’ did not begin to be used in its restricted modern sense until the middle of the last century.

As scientific discoveries accumulated it became increasingly difficult, and eventually impossible, for any one person to keep fully abreast of developments across the entire field. A point came when there was just too much information for a single brain to hold. Scientists themselves could no longer switch back and forth between disciplines as they used to do. They became specialists and during this century their specialisms have divided repeatedly. As a broadly educated person today, you may still have a general grasp of the basic principles of most of the specialisms, but not of the detail in which the research workers themselves are immersed. This is not your fault and you are not alone. Trapped inside their own specialisms, most research scientists find it difficult to communicate with those engaged in other research areas, even those bordering their own. No doubt you have heard the cliché defining a specialist as someone who knows more and more about less and less. We are in the middle of what journalists call an ‘information explosion’ and most of that information is being generated by scientists.

Clearly, the situation is unsatisfactory and there is a need to draw the specialisms into groups that will provide overarching views of broad topics. It should be possible, for example, to fit the work of the molecular biologist, extracting, cloning, and sequencing DNA, into some context that would relate it to the work of the taxonomist, and the work of both to that of the biochemist. What these disciplines share is their subject matter. All of them deal with living or once-living organisms. They deal with life and so these, as well as a whole range of related specialisms, have come to be grouped together as the life sciences. Similarly, geophysics, geochemistry, geomorphology, hydrology, mineralogy, pedology,

oceanography, climatology, meteorology, and other disciplines are now grouped as the earth sciences, because all of them deal with the physical and chemical nature of the planet Earth.

The third, and possibly broadest, of these groupings comprises the environmental sciences, sometimes known simply as ‘environmental science’. It embraces all those disciplines which are concerned with the physical, chemical, and biological surroundings in which organisms live. Obviously, environmental science draws heavily on aspects of the life and earth sciences, but there is some unavoidable overlap in all these groupings. Should palaeontology, for example, the study of past life, be regarded as a life science or, because its material is fossilized and derived from rocks, an earth science? It is both, but not necessarily at the same time. The palaeontologist may date a fossil and determine the conditions under which it was fossilized as an earth scientist, and as a life scientist reconstruct the organism as it appeared when it was alive and classify it. It is the direction of interest that defines the grouping.

Any study of the Earth and the life it supports must deal with process and change. The earth and life sciences also deal with process and change, but environmental science is especially concerned with changes wrought by human activities, and their immediate and long-term implications for the welfare of living organisms, including humans.

At this point, environmental science acquires political overtones and leads to controversy. If it suggests that a particular activity is harmful, then modification of that activity may require national legislation or an international treaty and, almost certainly, there will be an economic price that not everyone will have to pay or pay equally. We may all be environmental winners in the long term, but in the short term there will be financial losers and, not surprisingly, they will complain.

Over the last thirty years or so we have grown anxious about the condition of the natural environment and increasingly determined to minimize avoidable damage to it. In most countries, including the United States and European Union, there is now a legal requirement for those who propose any major development project to calculate its environmental consequences, and the resulting environmental impact assessment is taken into account when deciding whether to permit work to proceed. Certain activities are forbidden on environmental grounds, by granting protection to particular areas, although such protection is rarely absolute. It follows that people engaged in the construction, extractive, manufacturing, power-generating or power-distributing, agricultural, forestry, or distributive industries are increasingly expected to predict and take responsibility for the environmental effects of their activities. They should have at least a general understanding of environmental science and its application. For this reason, many courses in planning and industrial management now include an environmental science component.

This book provides an overview of the environmental sciences. As with all the broad scientific groupings, opinions differ as to which disciplines the term covers, but here the net is cast widely. All the topics it includes are generally accepted as environmental sciences. That said, the approach adopted in *Basics of Environmental Science* is not the only one feasible. In this rapidly developing field there is a variety of ideas about what should be included and emphasized and what constitutes an environmental scientist.

This opening chapter provides a general introduction to environmental science, its history, and its relationship to environmental campaigning. It is here that an important point is made about the overall subject and the content of the book: environmental science and ‘environmentalism’ are not at all the same thing. Environmental science deals with the way the natural world functions; environmentalism with such modifications of human behaviour as reformers think appropriate in the light of scientific findings. Environmentalists, therefore, are concerned with more than just science. As its title implies, *Basics of Environmental Science* is concerned mainly with the science.

The introduction is followed by four chapters, each of which deals with an aspect of the fundamental earth and life sciences on which environmental science is based, in each case emphasizing the importance of process and change and, where appropriate, relating the scientific description of what happens to its environmental implications and the possible consequences of perturbations to the system. The fifth and final chapter deals with environmental management, covering such matters as wildlife conservation, pest control, and the control of pollution.

You do not have to be a scientist to understand *Basics of Environmental Science*. Its language is simple, non-technical, and non-mathematical, but there are suggestions for further reading to guide those who wish to learn more. Nor do you have to read the book in order, from cover to cover. Dip into it in search of the information that interests you and you will find that each short block is quite self-contained.

It is the grouping of a range of disciplines into a general topic, such as environmental science, which makes it possible to provide a broad, non-technical introduction. The grouping is natural, in that the subjects it encompasses can be related to one another and clearly belong together, but it does not resolve the difficulty of scientific specialization. Indeed, it cannot, for the great volume of specialized information that made the grouping desirable still exists. Except in a rather vague sense, you cannot become an ‘environmental scientist’, any more than you could become a ‘life scientist’ or an ‘earth scientist’. Such imprecise labels have very little meaning. Were you to pursue a career in the environmental sciences you might become an ecologist, perhaps, or a geomorphologist, or a palaeoclimatologist. As a specialist you would contribute to our understanding of the environment, but by adding detailed information derived from your highly specialized research.

Environmental science exists most obviously as a body of knowledge in its own right when a team of specialists assembles to address a particular issue. The comprehensive study of an important estuary, for example, involves mapping the solid geology of the underlying rock, identifying the overlying sediment, measuring the flow and movement of water and the sediment it carries, tracing coastal currents and tidal flows, analysing the chemical composition of the water and monitoring changes in its distribution and temperature at different times and in different parts of the estuary, sampling and recording the species living in and adjacent to the estuary and measuring their productivity.¹ The task engages scientists from a wide range of disciplines, but their collaboration and final product identifies them all as ‘environmental scientists’, since their study supplies the factual basis against which future decisions can be made regarding the environmental desirability of industrial or other activities in or beside the estuary. Each is a specialist; together they are environmental scientists, and the bigger the scale of the issue they address the more disciplines that are likely to be involved. Studies of global climate change currently engage the attention of climatologists, palaeoclimatologists, glaciologists, atmospheric chemists, oceanographers, botanists, marine biologists, computer scientists, and many others, working in institutions all over the world.

You cannot hope to master the concepts and techniques of all these disciplines. No one could, and to that extent the old definition of an ‘educated person’ has had to be revised. Allowing that in the modern world no one ignorant of scientific concepts can lay serious claim to be well educated, today we might take it to mean someone possessing a general understanding of the scientific concepts from which the opinions they express are logically derived. In environmental matters these are the concepts underlying the environmental sciences. *Basics of Environmental Science* will introduce you to those concepts. If, then, you decide to become an environmental scientist the book may help you choose what kind of environmental scientist to be.

2 Environmental interactions, cycles, and systems

Inquisitive children sometimes ask whether the air they breathe was once breathed by a dinosaur. It may have been. The oxygen that provides the energy to power your body has been used many times by many different organisms, and the carbon, hydrogen, and other elements from which your body is made have passed through many other bodies during the almost four billion years that life has existed on our planet. All the materials found at the surface of the Earth, from the deepest ocean trenches to the top of the atmosphere, are engaged in cycles that move them from place to place. Even the solid rock beneath your feet moves, as mountains erode, sedimentary rocks are subducted into the Earth's mantle, and volcanic activity releases new igneous rock. There is nothing new or original in the idea of recycling!

The cycles proceed at widely differing rates and rates that vary from one part of the cycle to another. Cycling rates are usually measured as the time a molecule or particle remains in a particular part of the cycle. This is called its 'residence time' or 'removal time'. On average, a dust or smoke particle in the lower atmosphere (the troposphere) remains airborne for a matter of a few weeks at most before rain washes it to the surface, and a water molecule remains in the air for around 9 or 10 days. Material reaching the upper atmosphere (the stratosphere) resides there for much longer, sometimes for several years, and water that drains from the surface into ground water may remain there for up to 400 years, depending on the location.

Water that sinks to the bottom of the deep oceans eventually returns to the surface, but this takes very much longer than the removal of water molecules from the air. In the Pacific Ocean, for example, it takes 1000 to 1600 years for deep water to return to the surface and in the Atlantic and Indian Oceans it takes around 500 to 800 years (MARSHALL, 1979). This is relevant to concerns about the consequences of disposing industrial and low-level radioactive waste by sealing it in containers and dumping them in the deep oceans.

Those monitoring the movement of materials through the environment often make use of labelling, different labels being appropriate for different circumstances. In water, chemically inert dyes are often used. Certain chemicals will bond to particular substances. When samples are recovered, analysis reveals the presence or absence of the chemical label. Radioisotopes are also used. These consist of atoms chemically identical to all other atoms of the same element, but with a different mass, because of a difference in the number of neutrons in the atomic nucleus. Neutrons carry no charge and so take no part in chemical reactions, the chemical characteristics of an element being determined by the number of protons, with a positive charge, in its atomic nucleus.

You can work out the atmospheric residence time of solid particles by releasing particles labelled chemically or with radioisotopes and counting the time it takes for them to be washed back to the ground, although the resulting values are very approximate. Factory smoke belching forth on a rainy day may reach the ground within an hour or even less; the exhaust gases from an aircraft flying at high altitude will take much longer, because they are further from the ground to start with and in much drier air. It is worth remarking, however, that most of the gases and particles which pollute the air and can be harmful to health have very short atmospheric residence times. Sulphur dioxide, for example, which is corrosive and contributes to acid rain, is unlikely to remain in the air for longer than one month and may be washed to the surface within one minute of being released. The atmospheric residence time for water molecules is calculated from the rate at which surface water evaporates and returns as precipitation.

The deep oceans are much less accessible than the atmosphere, but water carries a natural label in the form of carbon-14(^{14}C). This forms in the atmosphere through the bombardment of nitrogen

(^{14}N) by cosmic radiation, but it is unstable and decays to the commoner ^{12}C at a steady rate. While water is exposed to the air, both ^{12}C and ^{14}C dissolve into it, but once isolated from the air the decay of ^{14}C means that the ratio of the two changes, ^{12}C increasing at the expense of ^{14}C . It is assumed that ^{14}C forms in the air at a constant rate, so the ratio of ^{12}C to ^{14}C is always the same and certain assumptions are made about the rate at which atmospheric carbon dioxide dissolves into sea water and the rate at which water rising from the depths mixes with surface water. Whether or not the initial assumptions are true, the older water is the less ^{14}C it will contain, and if the assumptions are true the age of the water can be calculated from its ^{14}C content in much the same way as organic materials are ^{14}C -dated.

Carbon, oxygen, and sulphur are among the elements living organisms use and they are being cycled constantly through air, water, and living cells. The other elements required as nutrients are also engaged in similar biogeochemical cycles. Taken together, all these cycles can be regarded as components of a very complex system functioning on a global scale. Used in this sense, the concept of a ‘system’ is derived from information theory and describes a set of components which interact to form a coherent, and often self-regulating, whole. Your body can be considered as a system in which each organ performs a particular function and the operation of all the organs is coordinated so that you exist as an individual who is more than the sum of the organs from which your body is made.

Biochemical cycles

The surface of the Earth can be considered as four distinct regions and because the planet is spherical each of them is also a sphere. The rocks forming the solid surface comprise the lithosphere, the oceans, lakes, rivers, and icecaps form the hydrosphere, the air constitutes the atmosphere, and the biosphere contains the entire community of living organisms.

Materials move cyclically among these spheres. They originate in the rocks (lithosphere) and are released by weathering or by volcanism. They enter water (hydrosphere) from where those serving as nutrients are taken up by plants and from there enter animals and other organisms (biosphere). From living organisms they may enter the air (atmosphere) or water (hydrosphere). Eventually they enter the oceans (hydrosphere), where they are taken up by marine organisms (biosphere). These return them to the air (atmosphere), from where they are washed to the ground by rain, thus returning to the land.

The idea that biogeochemical cycles are components of an overall system raises an obvious question: what drives this system? It used to be thought that the global system is purely mechanical, driven by physical forces, and, indeed, this is the way it can seem. Volcanoes, from which atmospheric gases and igneous rocks erupt, are purely physical phenomena. The movement of crustal plates, weathering of rocks, condensation of water vapour in cooling air to form clouds leading to precipitation—all these can be explained in purely physical terms and they carry with them the substances needed to sustain life. Organisms simply grab what they need as it passes, modifying their requirements and strategies for satisfying them as best they can when conditions change.

Yet this picture is not entirely satisfactory. Consider, for example, the way limestone and chalk rocks form. Carbon dioxide dissolves into raindrops, so rain is very weakly acid. As the rain water washes across rocks it reacts with calcium and silicon in them to form silicic acid and calcium bicarbonate, as separate calcium and bicarbonate ions. These are carried to the sea, where they react to form calcium carbonate, which is insoluble and slowly settles to the sea bed as a sediment that, in time, may be compressed until it becomes the carbonate rock we call limestone. It is an entirely inanimate process. Or is it? If you examine limestone closely you will see it contains vast numbers of shells, many of them minute and, of course, often crushed and deformed. These are of biological origin. Marine organisms ‘capture’ dissolved calcium and bicarbonate to ‘manufacture’ shells of calcium carbonate. When they die the soft parts of their bodies decompose, but their insoluble shells sink to the sea bed. This appears to be the principal mechanism in the formation of carbonate rocks and it has occurred on a truly vast scale, for limestones and chalks are among the commonest of all sedimentary rocks. The famous White Cliffs of Dover are made from the shells of once-living marine organisms, now crushed, most of them beyond individual recognition.

Here, then, is one major cycle in which the biological phase is of such importance that we may well conclude that the cycle is biologically driven, and its role extends further than the production of rock. The conversion of soluble bicarbonate into insoluble calcium carbonate removes carbon, as carbon dioxide, from the atmosphere and isolates it. Eventually crustal movements may return the rock to the surface, from where weathering returns it to the sea, but its carbon is in a chemically stable form. Other sedimentary rock on the ocean floor is subducted into the mantle. From there its carbon is returned to the air, being released volcanically, but the cycle must be measured in many millions of years. For all practical purposes, most of the carbon is stored fairly permanently. As the newspapers constantly remind us, carbon dioxide is a ‘greenhouse gas’, one of a number of gases present in the atmosphere that are transparent to incoming, short-wave solar radiation, but partially opaque to long-wave radiation emitted from the Earth’s surface when the Sun has warmed it. These gases trap outgoing heat and so maintain a temperature at the surface markedly higher than it would be were they absent. Since the Earth formed, some 4.6 billion years ago, the Sun has grown hotter by an estimated 25 to 30 per cent, and the removal of carbon dioxide from the air, to a significant extent as a result of biological activity, has helped prevent surface temperatures rising to intolerable levels.

Gaia

A hypothesis, proposed principally by James Lovelock, that all the Earth’s biogeochemical cycles are biologically driven and that on any planet which supports life conditions favourable to life are maintained biologically. Lovelock came to this conclusion as a result of his participation in the preparations for the explorations of the Moon and Mars. One object of the Mars programme was to seek signs of life on the planet. Martian organisms, should they exist, might well be so different from organisms on Earth as to make them difficult to recognize as being alive at all. Lovelock reasoned that the one trait all living organisms share is their modification of the environment. This occurs when they take materials from the environment to provide them with energy and structural materials, and discharge their wastes into the environment. He argued that it should be possible to detect the presence of life by an environment, especially an atmosphere, that was far from chemical equilibrium. Earth has such an atmosphere, with anomalously

large amounts of nitrogen and oxygen, as well as methane, which cannot survive for long in the presence of oxygen. It then occurred to him that the environmental modifications made and sustained by living organisms actually produced and maintained chemical and physical conditions optimum for those organisms themselves. In other words, the organisms produce an environment which suits them and then ‘manage’ the planet in ways that maintain those conditions.

Does this suggest that our climate is moderated, or even controlled, by biological manipulation? Certainly this is the view of James Lovelock, whose Gaia hypothesis takes the idea much further, suggesting that the Earth may be regarded as, or perhaps really *is*, a single living organism. It was this idea of a ‘living planet’ that he came to call ‘Gaia’ (LOVELOCK, 1979).

His hypothesis has aroused considerable interest, but Gaia remains controversial and there are serious objections to it. Expressed in its most extreme form, which is that almost all surface processes are biologically driven, it appears circular, with an explanation for everything, as when the existence of Gaia is introduced to explain the hospitable environment and the hospitable environment proves the existence of Gaia (JOSEPH, 1990). On the other hand, the more moderate version, which emphasizes the biological component of biogeochemical cycles more strongly than most traditional accounts, commands respect and promises to be useful in interpreting environmental phenomena, although not all scientists would associate this with the name ‘Gaia’ (WESTBROEK, 1992). It has been found, for example, that the growth of marine plankton can be stimulated by augmenting the supply of iron, an essential and, for them, limiting nutrient, with implications for the rate at which carbon dioxide is transferred from the atmosphere to the oceans and, therefore, for possible climate change (DE BAAR *ET AL.*, 1995).

Authorities differ in the importance they allot to the role of the biota (the total of all living organisms in the world or some defined part of it) in driving the biogeochemical cycles, but all agree that it is great, and it is self-evident that the constituents of the biota shape their environment to a considerable extent. Grasslands are maintained by grazing herbivores, which destroy seedlings by eating or trampling them, so preventing the establishment of trees, and over-grazing can reduce semi-arid land to desert. The presence of gaseous oxygen in the atmosphere is believed to result from photosynthesis.

We alter the environment by the mere fact of our existence. By eating, excreting, and breathing we interact chemically with our surroundings and thereby change them. We take and use materials, moving them from place to place and altering their form. Thus we subtly modify environmental conditions in ways that favour some species above others. In our concern that our environmental modifications are now proceeding on such a scale as to be unduly harmful to other species and possibly ourselves, we should not forget that in this respect we differ from other species only in degree. All living things alter their surroundings, through their participation in the cycles that together comprise the system which is the dynamic Earth.

3 Ecology and environmentalism

Our concern over the condition of the natural environment has led to the introduction of a new concept, of ‘environmental quality’, which can be measured against defined parameters. To give one example, if the air contains more than 0.1 parts per million (ppm) of nitrogen dioxide (NO_2) or sulphur dioxide

(SO₂) persons with respiratory complaints may experience breathing difficulties, and if it contains more than about 2.5 ppm of NO₂ or 5.0 ppm of SO₂ healthy persons may also be affected (KUPCHELLA AND HYLAND, 1986). These are quantities that can be monitored, and there are many more. It is also possible, though much more difficult, to determine the quality of a natural habitat in terms of the species it supports and to measure any deterioration as the loss of species.

These are matters that can be evaluated scientifically, in so far as they can be measured, but not everything can be measured so easily. We know, for example, that in many parts of the tropics primary forests are being cleared, but although satellites monitor the affected areas it is difficult to form accurate estimates of the rate at which clearance is proceeding, mainly because different people classify forests in different ways and draw different boundaries to them. The United Nations Environment Programme (UNEP) has pointed out that between 1923 and 1985 there were at least 23 separate estimates of the total area of closed forest in the world, ranging from 23.9 to 60.5 million km². The estimate UNEP prefers suggests that in pre-agricultural times there was a total of 12.77 million km² of tropical closed forest and that by 1970 this had been reduced by 0.48 per cent, to 12.29 million km², and that the total area of forests of all kinds declined by 7.01 per cent, from 46.28 to 39.27 million km², over the same period (TOLBA *ET AL.*, 1992). Edward O. Wilson, on the other hand, has written that in 1989 the total area of *rain* forests was decreasing by 1.8 per cent a year (WILSON, 1992). (A rain forest is one in which the annual rainfall exceeds 2540 mm; most occur in the tropics, but there are also temperate rain forests.) Similar differences occur in estimates of the extent of land degradation through erosion and the spread of deserts (called ‘desertification’). Before we can devise appropriate responses to these examples of environmental deterioration we have to find some way of reconciling the varying estimates of their extent. After all, it is impossible to address a problem unless we can agree on its extent.

Even when quantities can be measured with reasonable precision controversy may attend interpretations of the measurements. We can know the concentration of each substance present in air, water, soil, or food in a particular place at a particular time. If certain of those substances are not ordinarily present and could be harmful to living organisms we can call them ‘pollutants’, and if they have been introduced as a consequence of human activities, rather than as a result of a natural process such as volcanism, we can seek to prevent further introduction of them in the future. This may seem simple enough, but remember that someone has to pay for the measurement: workers need wages, and equipment and materials must be bought. Reducing pollution is usually inconvenient and costly, so before taking action, again we need to determine the seriousness of the problem. The mere presence of a pollutant does not imply harm, even when the pollutant is known to be toxic. Injury will occur only if susceptible organisms are exposed to more than a threshold dose, and where large numbers of very different species of plants, animals, and microorganisms are present this threshold is not easily calculated.

Nor is it easy to calculate thresholds for human exposure, because only large populations can be used for the epidemiological studies that will demonstrate effects, and small changes cannot always be separated statistically from natural fluctuations. (Epidemiology is the study of the incidence, distribution, and control of illness in a human population.) It has been estimated that over several decades the 1986 accident at the Chernobyl nuclear reactor may lead to a 0.03 per cent increase in radiation-induced cancer deaths in the former Soviet Union and a 0.01 per cent increase in the world as a whole, increases that will not be detectable against the natural variations in the incidence of cancer from year to year (ALLABY, 1995).

Where there is doubt, prudence may suggest we set thresholds very low, and in practice this is what happens. With certain pesticide residues in food, for example, the EU operates a standard of ‘surrogate zero’ by setting limits lower than the minimum quantity that can be detected.

Where the statistical evaluation of risk is unavoidably imprecise yet remedial action seems intuitively desirable, decisions cannot be based solely on scientific evidence and are bound to be more or less controversial. Since decisions of any kind are necessarily political, and will be argued this way and that, people will take sides and issues will tend to become polarized.

At this point, environmental science gives way to environmental campaigning, or environmentalism, and political campaigns are managed by those activists best able to publicize their opinion. In their efforts to attract public attention and support, spokespersons are likely to be drawn into oversimplifying complex, technical issues which, indeed, they may not fully understand, and to exaggerate hazards for the sake of dramatic effect.

Environmental science has a long history and concern with the condition of the environment has been expressed at intervals over many centuries, but the modern environmental movement emerged during the 1960s, first in the United States and Britain. The publication of *Silent Spring* in 1962 in the United States and 1963 in Britain provided a powerful stimulus to popular environmental concern and may have marked the origin of the modern movement. This was the book in which Rachel Carson mounted a strong attack on the way agricultural insecticides were being used in North America. The dire consequences of which she warned were essentially ecological: she maintained that the indiscriminate poisoning of insects by non-selective compounds was capable of disrupting food chains, the sequences of animals feeding on one another as, for example, insects ? blackbirds ? sparrowhawks. The 'silent spring' of her title referred to the absence of birds, killed by poisons accumulated through feeding on poisoned insects, but the 'fable' with which the book begins also describes the deaths of farm livestock and humans. The catastrophe was ecological and so the word 'ecology' acquired a political connotation. A magazine devoted to environmental campaigning, founded in 1970, was (and still is) called *The Ecologist*.

Ecology is a scientific discipline devoted to the study of relationships among members of living communities and between those communities and their abiotic environment. Intrinsically it has little to do with campaigning for the preservation of environmental quality, although individual ecologists often contribute their professional expertise to such campaigns and, of course, their services are sought whenever the environmental consequences of a proposed change in land use are assessed.

To some non-scientists, however, 'ecology' suggests a kind of stability, a so-called 'balance of nature' that may have existed in the past but that we have perturbed. This essentially metaphysical concept is often manifested as an advocacy for ways of life that are held to be more harmonious or, in the sense in which the word is now being used, 'ecological'. The idea is clearly romantic and supported by a somewhat selective view of history, but it has proved powerfully attractive. In her very detailed study of it Meredith Veldman, a historian at Louisiana State University, locates the development of environmentalism in Britain firmly in a long tradition of romantic protest that also includes the fiction of J.R.R.Tolkien and the Campaign for Nuclear Disarmament (VELDMANN, 1994).

'Ecology', then, is at one and the same time a scientific discipline and a political, at times almost religious, philosophy which inspires a popular movement and 'green' political parties in many countries. As a philosophy, it no longer demands piecemeal reform to achieve environmental amelioration, but calls for the radical restructuring of society and its economic base. The two meanings attached to the word are now quite distinct and it is important not to confuse them. When people say a particular activity or way of life is 'ecologically sound' they are making a political statement, not a scientific one, even though they may be correct in supposing the behaviour they approve to have less adverse effect on human health or the welfare of other species than its alternatives. 'Ecologically sound' implies a moral judgement that has no place in scientific argument; to a scientist the phrase is meaningless.

This is not to denigrate those who use the word ‘ecology’ in one sense or the other, simply to point out that the meanings are distinct and our attitudes to the environment are shaped by historical, social, and economic forces. They are not derived wholly from a scientific description of the environment or understanding of how it works. The nuclear power industry, for example, is opposed on ecological grounds, but there is no evidence that it has ever caused the slightest injury to non-humans, apart from vegetation around the Chernobyl complex following the accident there, and its adverse effects on human health are extremely small, especially when compared with those resulting from other methods of power generation; indeed, it is extremely unlikely that the correct routine operation of a nuclear power plant has any harmful effect at all, on humans or non-humans.² The anti-nuclear wing of the environmental movement is highly influential and has done much to erode public confidence in the industry, but whether this is environmentally beneficial is open to debate, to say the least. In contrast, on those occasions when scientists and campaigners collaborate, say in devising (scientifically) the best way to manage an area in order to maximize its value as natural habitat then campaigning (politically) to have the area protected from inappropriate development, they can achieve their useful and practicable goal. While it is certainly true that some ecological (i.e. environmentalist) campaigns owe little to ecology (the science), others, though not necessarily the most populist, are scientifically well informed. It is also true that if we confine our interest to the acquisition of an abstract understanding of the way the world is, that understanding will be of limited practical value. If damage to the environment is to be avoided or past damage remedied, scientific understanding must be applied and this is possible only through political processes.

This book will introduce you to the environmental sciences, of which ecology is one and, therefore, the word ‘ecology’ will henceforth be used only in its scientific sense. When issues of concern to environmentalists are discussed, as obviously they must be, they will be evaluated scientifically rather than politically. If your knowledge of environmental matters until now has been derived principally from campaigning literature, you may find the scientific accounts describe a world that is far more complex than you may have supposed and about which rather less is known than the campaigners sometimes imply. You should not be disheartened, for that is the way it is, and much remains to be discovered—perhaps one day, by you.

4 History of environmental science

By the time their civilization reached its peak in the Fifth Dynasty (after about 2480 BC) the ancient Egyptians seem to have become happy people. According to accounts described by the late Joseph Campbell (CAMPBELL, 1962), a leading authority on the ways people have seen themselves and the world around them, they had a joyful, outward-looking view of the world around them. True, they were somewhat preoccupied with the after-life, but that was seen pretty much as a continuation of their present lives and was celebrated in some of the most beautiful art and magnificent architecture the world has ever seen. Their pharaoh was described as ‘good’ rather than ‘great’ and the land he ruled was paradise, mythologically and to some extent literally. Life was very predictable and secure. Each year, the appearance of Sirius, the star of Isis, on the horizon at dawn heralded the flooding of the Nile. The reliable flood brought water and silt to enrich the cultivated land and guarantee the bountiful harvest that would follow. No doubt the work was hard, as it always is, but there was ample time for festivals and celebrations.

The Egyptians did not develop what we would recognize as science. Their view of the world was mythological and magical. Nevertheless, they did have a view of the world and a practical knowledge of those aspects of it that mattered to them. They knew much about agriculture, plants and animals,

water and how to use it effectively, they knew how to make bricks and were expert in the use of stone. People have always constructed mental frameworks to describe and explain the world around them. Not all were as positive as that of the Egyptians, but humans have an inherent need to understand, to make sense of their surroundings and locate themselves in them.

If we are to understand the world about us we must discover an order underlying phenomena or, failing that, impose one. Only then can we categorize things and so bring coherence to what would otherwise be chaotic. Most early attempts at classification were based on a mythological world-view. The anthropologist Mary Douglas has suggested, for example, that the biblical distinction between ‘clean’ and ‘unclean’ animals arose because Hebrew priests believed that sheep and goats, both ruminant animals with cloven hoofs, fitted into what they supposed to be the divine scheme, but pigs did not, because they have cloven hoofs but are not ruminants (BOWLER, 1992, pp. 11–12).

Science, in those days called ‘philosophy’ ('love of wisdom'), began with Thales (c. 640–546 BC), who lived in the Greek trading town of Miletus on the Aegean coast of what is now Turkey. He and his followers became known as the Ionian or Milesian school and the radical idea they introduced was that phenomena could be discussed rationally. That is to say, they suggested the mythical accounts of creation could be tested and rational explanations proposed for the order underlying the constant change we see everywhere. It is this critical attitude, allowing all ideas to be challenged by rational argument based on evidence and weaker theories to be replaced by stronger ones, which distinguishes science from non-science and pseudoscience. It originated only once; other civilizations developed considerable technological skills, but it was only among the Greeks living on the shores of Asia Minor that the modern concept of a ‘scientific approach’ emerged. All our science is descended from that beginning, and it began with environmental science. The Greek development reached its peak with the Academy, founded by Plato (429–347 BC), a student of Socrates, and the Lyceum, founded by Plato’s disciple Aristotle (384–322 BC). Aristotle wrote extensively on natural history. His studies of more than 500 species of animals included accurate descriptions, clearly based on personal observation, that were not confirmed until many centuries later. He recorded, for example, the reproduction of dogfish and the mating of squid and octopus. He also wrote about the weather in a book called *Meteorologica* ('discourse on atmospheric phenomena'), from which we derive our word ‘meteorology’.

Roman thinkers continued the Greek tradition, Pliny the Elder (c. AD 23–79) being the best-known Roman naturalist. His *Natural History*, covering what are now recognized as botany, zoology, agriculture, geography, geology, and a range of other topics, was based on fact, although he mingled records of his own observations with myths and fantastic travellers’ tales. Muslim scholars translated the Greek and Latin texts into Arabic, but it was not until the thirteenth century that they became generally available in Europe, as Latin translations from the Arabic.

Throughout this long history the central purpose of the enterprise has survived. There have been digressions, confusions, theories that led into blind alleys, but always the principal aim has been to replace mythical explanations with rational ones. Since myth is very often enshrined in religious texts, it may seem that the scientific agenda is essentially atheistic. Indeed, it has been so at times and in respect of some religions, and to this day scientists are often accused of atheism, but most modern thinkers regard the conflict as much more apparent than real. The writings of the Arabian physician Avicenna (979–1037) and philosopher Averroës (1126–98) kept classical ideas current in the Muslim world, where they were accommodated quite comfortably by Islam, and St Thomas Aquinas (c. 1224–74) used the natural order revealed by Aristotle as a proof of the existence of God, thus permitting science and religion to coexist in Christendom.

This is not to say that the dividing line between mythical and rational explanations was always clearly drawn, nor to deny that interpretations which undermined traditional beliefs sometimes generated fierce arguments. Scientists were often engaged in attempts to reconcile the two views and, then as now, scientific ideas could be attacked on essentially political grounds. In the years following the French Revolution, for example, conservatives in Britain used scriptural authority to justify the preservation of the social order. This led scientists supporting them to adapt the Neptunian theory of Abraham Gottlob Werner (1749–1817) so that it appeared to substantiate the story of Noah's flood. Werner proposed that the Earth was once covered entirely by an ocean, from which some rocks had crystallized and beneath which others had been deposited as sediment, the rocks being exposed through the gradual and continuing retreat of the waters. This obsession with the biblical flood continues in some English-speaking countries to the present day, from time to time with 'discoveries' of the remains of the Ark, although scientists elsewhere in Europe had ceased to take it seriously by the eighteenth century (BOWLER, 1992, pp. 129–130).

Much of the history of the environmental sciences revolves about the reconstruction of the history of the planet since it first formed. To a considerable extent, this reconstruction was based on interpretations of fossils, which were by no means always seen as the obvious remains of once-living organisms.³ Even when it became possible to use the fossils entrapped within them to arrange rock strata in a chronological sequence controversy continued over the assignment of dates to those strata, the mechanisms by which the rocks had assumed their present forms and distribution, and over the total age of the Earth itself. It was in his effort to solve this puzzle that in 1650 James Ussher (1581–1656), an Irish scholar and archbishop of Armagh, constructed what may have been the first theoretical model. Basing his chronology on the Old Testament, he concluded the Earth had been created in 4004 BC!

If the development of environmental science seems to have been dominated by the study of rocks and fossils, it is perhaps because elucidating the history of the planet was a necessary first step toward an understanding of its present condition and, in any case, the classification and distribution of plants and animals played a major role in it. The theory of evolution by natural selection was derived from Earth history, and Charles Darwin (1809–82) began his career as a geologist.

A unifying theme was supplied by Alexander von Humboldt (1769–1859). Mining engineer, geologist, geophysicist, meteorologist, and geographer, Humboldt spent the years from 1799 to 1804 exploring in tropical South America with his friend, the botanist Aimé Bonpland (1773–1858). His subsequent accounts greatly advanced knowledge of plant geography and his five-volume *Kosmos*, completed after his death, sought to demonstrate how physical, biological, and human activities combined to regulate the environment (BOWLER, 1992, pp. 204–211). This helped establish biogeography as a scientific discipline and applied a range of disciplines to the study of environments. Humboldt is also credited with having shifted science generally from its rather abstract preoccupations in the eighteenth century to its much greater reliance on observation and experiment characteristic of the nineteenth and twentieth.

Biogeography also fed back into the earth sciences. Plotting the distribution of present and extinct plants and animals played a major part in the development of the theory of continental drift by the German climatologist Alfred Wegener (1880–1930), who sought to explain the apparent fit between the coasts of widely separated continents, such as the west coast of Africa and east coast of South America, by postulating that the continents were once joined and have since drifted apart. He published this in 1915 as *Die Entstehung der Kontinente und Ozeane* (it did not appear in English until 1924, as *The Origin of Continents and Oceans*), which led in turn to the theory of sea-floor spreading, proposing that continental drift is driven by the expansion and contraction of the crust beneath the ocean floor, and then, in the 1960s and 1970s, to the unifying concept of plate tectonics.

Ecology grew partly out of theories of evolution that were being discussed during the eighteenth and nineteenth centuries. Darwinism is an ecological theory, after all, but this line of development branched, the other strand leading into German Romanticism. This was a very influential intellectual movement based on the idea that individual freedom and self-expression would bring people into close touch with a sublime reality surrounding us all and of which we long to become part. The discipline of ecology also originated in a quite different concept, that of the 'economy of nature'. This led to an idyllic view of nature as the harmonious product of all the countless interactions among living organisms and well able to supply human needs. Indeed, the view had strong links to natural theology, according to which God had so endowed all plants and animals with needs and the means to satisfy them as to guarantee that harmony among them would be preserved. This is the origin of the idea of a 'balance of nature' and, sentimental though it sounds, it taught that the interactions among organisms relate them in complex ways, and by early in the eighteenth century, long before the word 'ecology' was coined (by Ernst Haeckel (1834–1919) in 1866), it had generated some ideas with a startlingly modern ring. The writer Richard Bradley (1688–1732), for example, noted that insect species tend to specialize in the plants on which they feed and he advised farmers not to kill birds in their fields, because the birds feed on insects that would otherwise damage crops.

Environmental science ranges so widely that much of the history of science is relevant to its own development. Even such apparently unrelated discoveries as the gas laws relate very directly to meteorology, climatology and, through them, to weather forecasting and considerations of possible climate change. Today, many disciplines contribute to environmental science and its practitioners are equipped with instruments and techniques that enable them to begin compiling an overall, coherent picture of the way the world functions. The picture remains far from complete, however, and we must be patient while we wait to discover whether some of what are popularly perceived as environmental problems are really so and, if they are, how best to address them.

5 Changing attitudes to the natural world

When Julius Caesar (100–44 BC) became emperor of Rome, in 47 BC, traffic congestion was one of the pressing domestic problems he faced. He solved it by banning wheeled traffic from the centre of Rome during daytime, with the predictable result that Romans were kept awake at night by the incessant rumbling of iron-shod wheels over cobblestones. Nevertheless, Claudio (10 BC–AD 54, reigned from 41) later extended the law to all the important towns of Italy, Marcus Aurelius (AD 121–80, reigned from 161) made it apply to every town in the empire, and Hadrian (AD 76–138, reigned from 117) tightened it by restricting the number of vehicles allowed to enter Rome even at night (MUMFORD, 1961). The problem then, as now, was that a high population density generates a high volume of traffic and no one considered the possibility of designing towns with lower population and housing densities, as an alternative to building more and bigger roads.

If environmental science has a long history, so do the environmental problems that concern us today. We tend to imagine that urban air pollution is a recent phenomenon, dating mainly from the period of rapid industrialization in Europe and North America that began in the late eighteenth century. Yet in 1306 a London manufacturer was tried and executed for disobeying a law forbidding the burning of coal in the city, and the first legislation aimed at reducing air pollution by curbing smoke emissions was enacted by Edward I in 1273. The early efforts were not particularly successful and they dealt only with smoke from the high-sulphur coal Londoners were importing by ship from north-east England and which was, therefore, known as 'sea coal'. A wide variety of industries contributed to the smells and dust and poured their effluents into the nearest river. The first attempts to reduce

pollution of the Thames date from the reign of Richard II (1367–1400, reigned from 1377). It was because of the smoke, however, that Elizabeth I refused to enter the city in 1578, and by 1700 the pollution was causing serious damage by killing vegetation, corroding buildings, and ruining clothes and soft furnishings in every town of any size (THOMAS, 1983). Indeed, the pall of smoke hanging over them was often the first indication approaching travellers had of towns.

Filthy it may have been, but ‘sea coal’ was convenient. It was a substitute for charcoal rather than wood, because of the high temperature at which it burned, and it was probably easier to obtain. If its use were to be curtailed, either manufacturing would suffer, with a consequent reduction in employment and prosperity, or charcoal would be used instead, in which case pollution might have been little reduced overall. Environmental protection always involves compromise between conflicting needs.

Much of the primary forest that once covered most of lowland Britain, which Oliver Rackham, possibly the leading authority on the history of British woodland, has called the ‘wildwood’, had been cleared by the time of the Norman invasion, in 1066, mainly to provide land on which to grow crops. It did not disappear, as some have suggested, to provide fuel for eighteenth-century iron foundries, or to supply timber to build ships. Paradoxically, the iron foundries probably increased the area of woodland, by relying for fuel on managed coppice from sources close at hand, and reports of a shortage of timber for shipbuilding had less to do with a lack of suitable trees than with the low prices the British Admiralty was prepared to pay (ALLABY, 1986, p. 110).

As early as the seventh century there were laws restricting the felling of trees and in royal forests a fence was erected around the stump of a felled tree to allow regeneration (ALLABY, 1986, p. 198). By the thirteenth century there were laws forbidding the felling of trees, clearing of woodland, and even the taking of dead wood, although they were seldom enforced, except as a means to raise revenue by fining an offender the value of the trees felled (RACKHAM, 1976).

For most of history, however, the conflict between farms and forests was resolved in favour of farms, although in England there is a possibility of confusion over the use of the word ‘forest’. Today, the word describes an extensive tract of land covered with trees growing closely together, sometimes intermingled with smaller areas of pasture. Under Norman law, however, it had a different meaning, derived from the Latin *foris*, meaning ‘outdoors’, and applied to land beyond the boundaries of the enclosed farmland or parklands and set aside for hunting. Much of this ‘forest’ belonged to the sovereign. Special laws applied to it and were administered by officers appointed for the purpose. It might or might not be tree-covered.

Forests were regarded as dark, forbidding places, the abode of dangerous wild animals and brigands.⁴ When Elizabethan writers used the word ‘wilderness’ they meant unmanaged forest, and in North America the earliest European settlers contrasted the vast forests they saw unfavourably with the cultivated fields they hoped to establish. Until modern times, famine was a real possibility and the neater the fields, the fewer the weeds in them, and the healthier the crops, the more reassuring the countryside appeared.

Mountains, upland moors, and wetlands were wastelands that could not be cultivated and they were no less alarming. In 1808, Arthur Young (1741–1820), an agricultural writer appointed secretary to the Board of Agriculture established by Prime Minister, William Pitt, in 1793, submitted a report on the enclosure of ‘waste’ land, arguing strongly in favour of their improvement by cultivation (YOUNG, 1808).

What we would understand today as the conservation of forest habitats and wildlife began quite early in the tropics, where it was a curious by-product of colonial expansion. This led government

agencies and private companies to employ scientists or, in the case of the British East India Company, surgeons, many of whom had time to spare and wide scientific interests. One of the earliest conservation experiments was begun in Mauritius in the middle of the eighteenth century by French reformers seeking to prevent further deforestation as part of their efforts to build a just society. Interestingly, they had perceived a relationship between deforestation and local climate change. In the British territories, scientists also noted this relationship. Forest reserves were established in Tobago in 1764 and St Vincent in 1791, and a law passed in French Mauritius in 1769 was designed to protect or restore forests, especially on hill slopes and near to open water. Plans for the planting and management of Indian forests began in 1847 (GROVE, 1992), the foresters being known as ‘conservators’, a title still used in Britain by the Forestry Commission.

At about the same time, Americans were also becoming aware of the need for conservation. George Perkins Marsh (1801–82), US ambassador to Italy from 1862 until his death, wrote *Man and Nature* while in Italy. Published in 1864, this book led to the establishment of forest reserves in the United States and other countries, but it also challenged the then accepted relationship between humans and the natural environment.⁵ In 1892, Warren Olney, John Muir, and William Keith founded the Sierra Club (www.sierraclub.org/index_right.htm), and the National Audubon Society (www.audubon.org/), named after John James Audubon (1785–1851),⁶ the renowned wildlife painter and conservationist, was founded shortly afterwards.

While ‘wilderness’ has always implied hostility (and nowadays the word is often applied to certain urban areas), to these early conservationists it also had another, quite different meaning. To them, and those who thought like them, the word suggested purity, freedom from human interference, and the place where humans may find spiritual renewal, although this idea was often combined, as it is still, with that of economic resources held in reserve until a use can be found for them. It is tempting to associate the spiritual view exclusively with European Romanticism, but it also occurs in non-European cultures and, even in Europe, there were a few writers who saw wilderness in this way prior to the eighteenth century.

Today, the love of wilderness and desire to protect it probably represents the majority view, at least in most industrialized societies. Similarly, most people recognize pollution as harmful and will support measures to reduce it, provided they are not too expensive or disruptive. As we have seen, however, these are far from being new ideas or new attitudes. They have emerged at various times in the past, then concern has waned. It may seem that public attitudes reflect some cyclical change, and this may be not far from the truth.

When the possibility of famine was real, the most beautiful landscape was one that was well and intensively farmed. When factory jobs were scarce and insecure, but for large numbers of people the only jobs available, smoking chimneys symbolized prosperity. No one could afford to care that the fumes were harmful, even that they were harmful to human health, for hunger and cold were still more harmful and more immediately so. When the first European colonists reached North America, they could make no living from the forest. They had to clear it to provide cultivable land, and they had to do so quickly. It was only the wealthy who had the leisure to contemplate wilderness and could afford to point out the dangers of pollution, with the risk that were their warnings heeded, factories might close.

Modern concerns continue to follow the cycle. The present wave of environmental concern began in Britain and the United States in the 1960s, at a time of rising prosperity. It continued into the 1970s and then, as economies began to falter and unemployment began to creep upwards, interest faded. It re-emerged in the 1980s, as economies seemed to revive, then waned again as recession began to bite hard. The fluctuations in public concern are recorded in the numbers of books on environmental topics published year by year. In the early 1970s vast numbers appeared, but far fewer books were

being published by the middle 1970s. More ‘green’ titles were issued in the early 1980s, but by the end of the decade large numbers of copies were being returned to publishers unsold and by the early 1990s most publishers would not accept books with titles suggesting anything remotely ‘ecological’, ‘environmental’, or ‘green’.

This should surprise no one. When times are hard people worry most about their jobs, their homes, and whether they will be able to feed their families. It is only when they have economic security that they feel able to relax sufficiently to turn their attention to other matters. The preservation of species or of a tranquil, attractive countryside in which to walk means little to the homeless teenager begging for food or the single mother whose child needs shoes.

It should surprise no one, but there is an important lesson to be learned from it. All governments now accept the need for environmental reform, but a perceptual gulf exists between rich and poor which parallels that between rich and poor within nations. In poor countries facing high levels of infant mortality and chronic shortages of supplies necessary for the provision of health care, housing, and education, the most pressing needs relate to the provision of employment and industrialization based, so far as possible, on the exploitation of indigenous resources. Environmental hazards seem less urgent, and efforts by the rich to persuade the poor to move them higher up the international agenda can be seen as attempts to increase development costs and so perpetuate economic inequality. It is well to remember that environmental issues that seem self-evidently urgent to Europeans and North Americans may not seem so to everyone.

End of chapter summary

Like the life and earth sciences, the environmental sciences comprise elements taken from many other disciplines. The all-round environmental scientist must be part biologist, part ecologist, part toxicologist, part pedologist (soil scientist), part geomorphologist, part limnologist (student of freshwater systems), and part meteorologist, as well as being familiar with ideas taken from many other disciplines. In helping to resolve the disputes that frequently arise over conservation and land use, the environmental scientist must also possess tact and political skill.

It is important that the environmental scientist remain a scientist. Environmentalism consists in campaigning and proselytising in pursuit of essentially ideological objectives. It is not necessarily based on scientific assessment, nor should it be, because its appeal is primarily moral. This means there is a clear difference between environmental science and environmentalism and the two should not be confused.

End of chapter points for discussion

- To what extent is our attitude to the natural world linked to our level of economic prosperity?
- To what extent are environmentalist campaigns informed by science?
- What does Gaia hypothesis assert?

See also

Formation of the Earth (section 6)
Weathering (section 8)
Coasts, estuaries, sea levels (section 10)
Greenhouse effect (section 13)
Evolution and structure of the atmosphere (section 14)
Dating methods (section 19)

Climate change (section 20)
Fresh water (section 22)
Biogeography (section 32)
Nutrient cycles (section 33)
Ecology (sections 35 – 44)
Limits of tolerance (section 44)
Evolution (section 45)
Transnational pollution (section 62)

Further reading

The Ages of Gaia. James Lovelock. 1988. Oxford University Press, Oxford. Provides a general, non-technical introduction to and description of the Gaia theory, written by the scientist who first proposed it.

Fantasy, the Bomb, and the Greening of Britain. Meredith Veldman. 1994. Cambridge University Press, New York. The rise of the environmental movement in Britain, seen in the context of a long history of romantic protest; written by a historian.

The Fontana History of the Environmental Sciences. Peter J.Bowler. 1992. Fontana Press (HarperCollins), London. Probably the most authoritative account of the subject; written in non-technical language.

Gaia: The Growth of an Idea. Lawrence E.Joseph. 1990. Arkana (Penguin Books), London. Also provides an account of Gaia theory, simply written by a journalist, but includes objections to the idea and difficulties it raises.

Man and the Natural World. Keith Thomas. 1983. Penguin Books, London. A comprehensive account of changing attitudes in Britain between 1500 and 1800.

Thinking Green: An Anthology of Essential Ecological Writing. Michael Allaby (ed.). 1989. Barrie and Jenkins, London. A selection of excerpts from some of the most influential writing on environmental topics.

Notes

- 1 See, for example, *The Clyde Estuary and Firth: an assessment of present knowledge*, compiled by members of the Clyde Study Group (1974), NERC Publications Series C No. 11.
- 2 For a useful discussion of this topic, see KUPCHELLA AND HYLAND, pp. 160–162.
- 3 See Rudwick, Martin J.S. 1976. *The Meaning of Fossils: Episodes in the history of palaeontology*. Univ. of Chicago Press, Chicago.
- 4 This theme is explored in Allaby, Michael, 1999, *Ecosystems: Temperate Forests*. Fitzroy Dearborn, London, pp. 146–149.
- 5 For an excerpt from this book see Allaby, Michael (ed.). 1989. *Thinking Green: An anthology of essential ecological writing*. Barrie and Jenkins, London, pp. 61–68.
- 6 For a brief biography of Audubon, see www.audubon.org/nas/jja.html. There is also an excellent essay about his relationship with Darwin in Weissmann, Gerald, 1998, *Darwin's Audubon*. Plenum Press, New York, pp. 9–24.

References

- Allaby, Michael. 1986. *The Woodland Trust Book of British Woodlands*. David and Charles, Newton Abbot. 1995. *Facing the Future*. Bloomsbury, London, p. 197.
- Bowler, Peter J. 1992. *The Fontana History of the Environmental Sciences*. FontanaPress (HarperCollins), London.
- Campbell, Joseph. 1962. *The Masks of God: Oriental Mythology*. Arkana (Penguin Books), London, pp. 95–98.
- de Baar, Hein J.W., de Jong, Jeroen T.M., Bakker, Dorothee C.E., Loscher, Bettina M., Veth, Cornelius, Bathmann, Uli, and Smetacek, Victor. 1995. ‘Importance of iron for plankton blooms and carbon dioxide drawdown in the Southern Ocean’. *Nature*, 373, 412–415.

- Grove, Richard H. 1992. 'Origins of Western environmentalism'. *Scientific American*, July 1992. pp. 22–27.
- Joseph, Lawrence E. 1990. *Gaia: The growth of an idea*. Arkana (Penguin Books), London. Ch. IV.
- Kupchella, Charles E. and Hyland, Margaret C. 1986. *Environmental Science*. Allyn and Bacon, Needham Heights, Mass. 2nd edition, pp. 316–317.
- Lovelock, James. 1979. *Gaia: A new look at life on Earth*. OUP, Oxford. 1988. *The Ages of Gaia*. OUP, Oxford.
- Marshall, N.B. 1979. *Developments in Deep-Sea Biology*. Blandford, Poole, Dorset, p.32–33.
- Mumford, Lewis. 1961. *The City in History*. Pelican Books, London, p. 254.
- Rackham, Oliver. 1976. *Trees and Woodland in the British Landscape*. J.M.Dent, London, p. 154. (A second, 1990, edition is now available.)
- Thomas, Keith. 1983. *Man and the Natural World*. Penguin Books, London, pp. 244–247.
- Tolba, Mostafa K. and El-Kholy, Osama A. 1992. *The World Environment 1972–1992*. Chapman and Hall, London, on behalf of UNEP. pp. 160–162.
- Veldman, Meredith. 1994. *Fantasy, the Bomb, and the Greening of Britain*. Cambridge University Press, New York.
- Westbroek, Peter. 1992. *Life as a Geological Force*. W.W.Norton, New York.
- Wilson, Edward O. 1992. *The Diversity of Life*. Penguin Books, London, p. 264.
- Young, Arthur. 1808. *General Report on Enclosures*. Republished, 1971, by Augustus M.Kelley, New York.

2 Earth Sciences

When you have read this chapter you will have been introduced to:

- the formation and structure of the Earth
- rocks, minerals, and geologic structures
- weathering
- how landforms evolve
- coasts, estuaries, and changing sea levels
- solar energy
- albedo and heat capacity
- the greenhouse effect
- evolution, composition, and structure of the atmosphere
- general circulation of the atmosphere
- ocean currents and gyres
- weather and climate
- ice ages and interglacials
- climate change
- climatic regions and plants

6 Formation and structure of the Earth

Among the nine planets in the solar system, Earth is the only one which is known to support life. All the materials we use are taken from the Earth and it supplies us with everything we eat and drink. It receives energy from the Sun, which drives its climates and biological systems, but materially it is self-contained, apart from the dust particles and occasional meteorites that reach it from space (ADAMS, 1977, pp. 35–36). These may amount to 10000 tonnes a year, but most are vaporized by the heat of friction as they enter the upper atmosphere and we see them as ‘shooting stars’. At the most fundamental level, the Earth is our environment.

The oldest rocks, found on the Moon, are about 4.6 billion years old and this is generally accepted to be the approximate age of the Earth and the solar system generally. There are several rival theories describing the process by which the solar system may have formed.¹ The most widely accepted theory, first proposed in 1644 by René Descartes (1596–1650), proposes that the system formed from the condensation of a cloud of gas and dust, called the ‘primitive solar nebula’ (PSN). It is now thought this cloud may have been perturbed by material from a supernova explosion. Fusion processes within stars convert hydrogen to helium and in larger stars go on to form all the heavier elements up to iron. Elements heavier than iron can be produced only under the extreme conditions of the supernova explosion of a very massive star, and the presence of such elements (including zinc, gold, mercury, and uranium) on Earth indicates a supernova source.

As the cloud condensed, its mass was greatest near the centre. This concentration of matter comprised the Sun, the planets forming from the remaining material in a disc surrounding the star, and the whole system

rotated. The inner planets formed by accretion. Small particles moved close to one another, were drawn together by their mutual gravitational attraction, and as their masses increased they gathered more particles and continued to grow. At some point it is believed that a collision between the proto-Earth and a very large body disrupted the planet, the material re-forming as two bodies rather than one: the Earth-Moon system. This explains why the Earth and Moon are considered to be of the same age and, therefore, why lunar rocks 4.6 billion years old are held to be of about the age of the Earth and Moon.

The material of Earth became arranged in discrete layers, like the skins of an onion. If accretion was a slow process compared to the rate at which the PSN cooled, the densest material may have arrived first, followed by progressively less dense material, in which case the layered structure has existed from the start and would not have been altered by melting due to the gravitational energy released as heat by successive impacts. This model is called ‘heterogeneous accretion’. If material arrived quickly in relation to the rate of PSN cooling, then it would have comprised the whole range of densities. As the planet cooled from the subsequent melting, denser material would have gravitated to the centre and progressively less dense material settled in layers above it. This model is called ‘homogeneous accretion’ (ALLABY AND ALLABY, 1999).

As it exists today, the Earth has a mean radius of 6371 km, equatorial circumference of 40077 km, polar circumference of 40009 km, total mass of 5976×10^{24} g, and mean density of 5.517 g. cm⁻³. Of its surface area, 149×10^6 km² (29.22 per cent) is land, 15.6×10^6 km² glaciers and ice sheets, and 361×10^6 km² oceans and seas (HOLMES, 1965, ch. II). Land and oceans are not distributed evenly. There is much more land in the northern hemisphere than in the southern, but at the poles the positions are reversed: Antarctica is a large continent, but there is little land within the Arctic Circle.

At its centre, the Earth has a solid inner core, 1370 km in radius, made from iron with some nickel (see Figure 2.1). This is surrounded by an outer core, about 2000 km thick, also of iron with nickel, but liquid, although of very high density. Movement in the outer core acts like a self-exciting dynamo and generates the Earth’s magnetic field, which deflects charged particles reaching the Earth from space. Outside the outer core, the mantle, made from dense but somewhat plastic rock, is about 2900 km thick, and at the surface there is a thin crust of solid rock, about 6 km thick beneath the oceans and 35 km thick (but less dense) beneath the continents.

Miners observed long ago that the deeper their galleries the warmer they found it to work in them. Surface rocks are cool, but below the surface the temperature increases with depth. This is called the ‘geothermal gradient’. A little of the Earth’s internal heat remains from the time of the planet’s formation, but almost all of it is due to the decay of the radioactive elements that are distributed widely throughout the mantle and crustal rocks. The value of the geothermal gradient varies widely from place to place, mostly between 20 and 40°C for every kilometre of depth, but in some places,

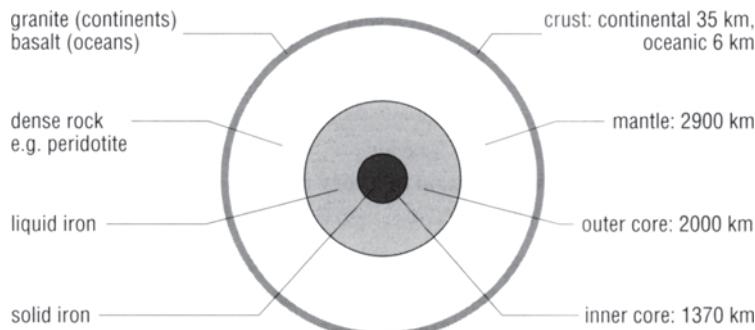


Figure 2.1 Structure of the Earth (not to scale)

such as Ontario, Canada, and the Transvaal, South Africa, it is no more than 9 or 10°C per kilometre (HOLMES, 1965, ch. XXVIII, p. 995). Because of the low thermal conductivity of rock, very little of this heat reaches the surface and it has no effect on the present climate.

Where the gradient is anomalously high, however, it can be exploited as a source of geothermal energy. In volcanic regions, such as New Zealand, Japan, Iceland, and Italy, water heated below ground may erupt at the surface as geysers, hot springs, or boiling mud. More often it fails to reach the surface and is trapped at depth, heated by the surrounding rock. A borehole drilled into such a reservoir may bring hot water to the surface where it can be used. In some places a body of dry subsurface rock is much hotter than its surroundings. In principle this can also be exploited, although experimental drilling, for example some years ago in Cornwall, Britain, has found the resulting energy rather costly. The technique is to drill two boreholes and detonate explosive charges at the bottom, to fracture the rock between them and so open channels through it. Cold water is then pumped at pressure down one borehole; it passes through the hot rock and returns to the surface through the other borehole as hot water.

This exploitation of geothermal energy is not necessarily clean. Substances from the rock dissolve into the water as it passes, so it returns to the surface enriched with compounds some of which are toxic. The solution is often corrosive and must be kept isolated from the environment and its heat transferred by heat exchangers. Nor is the energy renewable. Removal of heat from the rock cools it faster than it is warmed by radioactive decay, so eventually its temperature is too low for it to be of further use. Similarly, the abstraction of subsurface hot water depletes, and eventually empties, the reservoir.

Although subsurface heat has no direct climatic effect, there is a sense in which it does have an indirect one. Material in the mantle is somewhat plastic. Slow-moving convection currents within the mantle carry sections of the crustal rocks above them, so that over very long time-scales the crustal material is constantly being rearranged.² On Earth, but possibly on no other solar-system planet, the crust consists of blocks, called ‘plates’, which move in relation to one another. The theory describing the process is known as ‘plate tectonics’ (GRAHAM, 1981). At present there are seven large plates, a number of smaller ones, and a still larger number of ‘microplates’. The boundaries (called ‘margins’) between plates can be constructive, destructive, or conservative. At constructive margins two plates are moving apart and new material emerges from the mantle and cools as crustal rock to fill the gap, marked by a ridge. There are ridges near the centres of all the world’s oceans. Where plates move towards one another there is a destructive margin, marked by a trench where one plate sinks (is subducted) beneath the other. At conservative margins two plates move past one another in opposite directions (see Figure 2.2). There are also collision zones, where continents or island arcs have collided. In these, all the oceanic crust is believed to have been subducted into the mantle, leaving only continental crust. Such zones may be marked in various ways, one of which is the presence of mountains made from folded crustal rocks. An island arc is a series of volcanoes lying on the side of an ocean trench nearest to a continent. The volcanoes are due to the subduction of material.

Slowly but constantly the movement of plates redistributes the continents carried on them. A glance at a map shows the apparent fit between South America and Africa, but for 40 million years or more prior to the end of the Triassic Period, about 213 million years ago, all the continents were joined in a supercontinent, Pangaea, surrounded by a single world ocean, Panthalassa. Pangaea then broke into two continents, Laurasia in the north and Gondwana in the south, separated by the Tethys Sea, of which the present Mediterranean is the last remaining trace. The drift of continents in even earlier times has now been reconstructed, with the proposing of a supercontinent called Rodinia that existed about 750 million years ago (DALZIEL, 1995). The Atlantic Ocean opened about 200 million years ago and it is still growing wider by about 3–5 cm a year. A little more than 100 million years ago India

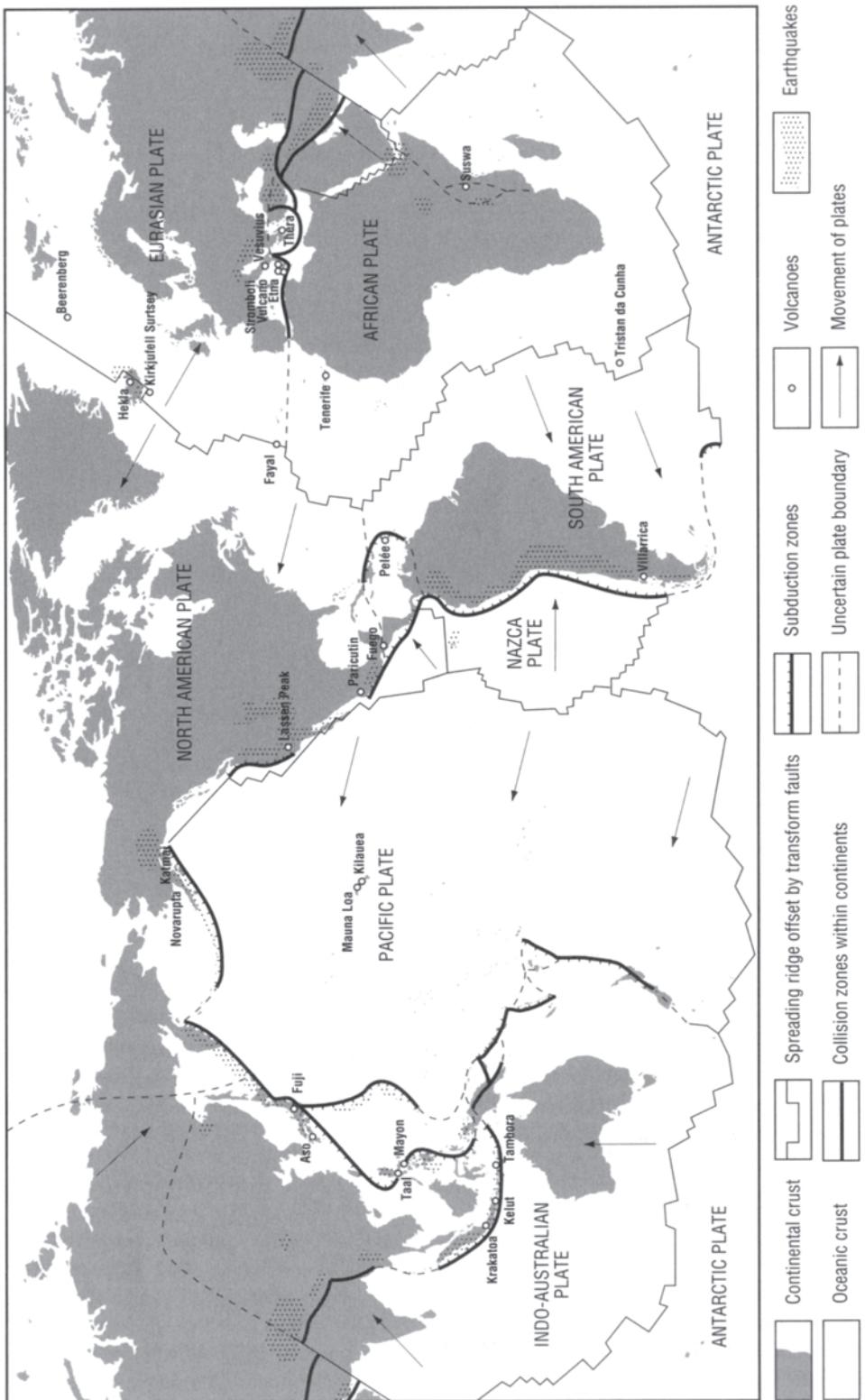


Figure 2.2 Plate structure of the Earth and seismically active zones

separated from Antarctica. The Indian plate began subducting beneath the Eurasian plate and as India moved north the collision, about 50 million years ago, raised the Himalayan mountain range. India is still moving into Asia at about 5 cm a year and the mountains are still growing higher (WINDLEY, 1984, pp. 161 and 310), although the situation is rather complicated. Rocks exposed at the surface are eroded by ice, wind, and rain, so mountains are gradually flattened. At the same time, the crumpling that produces mountains of this type increases the mass of rock, causing it to sink into the underlying mantle. This also reduces the height of large mountain ranges. It is possible, however, for the eroded material to lighten the mountains sufficiently to reduce the depression of the mantle, causing them to rise, and there is reason to suppose this is the case for the Himalayas (BURBANK, 1992). The Red Sea is opening and in time will become a new ocean between Africa and Arabia.

The distribution of land has a strong influence on climates. If there is land at one or other pole, ice sheets are more likely to form. The relative positions of continents modify ocean currents, which convey heat away from the equator, and the size of continents affects the climates of their interiors, because maritime air loses its moisture as it moves inland. The Asian monsoon is caused by pressure differences to the north and south of the Himalayas. In winter, subsiding air produces high pressure over the continent and offshore winds, with very dry conditions inland. The word ‘monsoon’ simply means ‘season’ (from the Arabic word for ‘season’, *mausim*) and this is the winter, or dry, monsoon. In summer, pressure falls as the land warms, the wind direction reverses, and warm, moist air flows across the ocean toward the continent, bringing heavy rain. This is the summer, wet monsoon. Plate tectonics exerts a very long-term influence, of course, and other factors modify climates in the shorter term, but the distribution of land and sea determines the overall types of climate the world is likely to have (HAMBREY AND HARLAND, 1981).

Plate tectonics affects the environment more immediately and more dramatically. The movement of plates causes earthquakes, because it tends to happen jerkily as accumulated stress is released, and is associated with volcanism due to weakening of the crust at plate margins. Earthquakes cause damage to physical structures, which is the direct cause of most injuries, and those which occur beneath the sea produce tsunami (www.geophys.washington.edu/tsunami/general/physics/physics.html). These are shock waves affecting the whole water column. No more than a metre high and with a wavelength of hundreds of kilometres, but travelling at more than 700 km h⁻¹, on reaching shallow water they rise to great height and destructive power (ALLABY, 1998, pp. 54–60).

If volcanic ash reaches the stratosphere it can cause climatic cooling, but volcanic eruptions are more usually associated with damage to human farms and dwellings. This arises partly because of the beneficial effect volcanoes can have. Volcanic ash and dust are often rich in minerals and rejuvenate depleted soils. Farmers can grow good crops on them, which is why there tend to be cultivated fields at the foot and even on the lower slopes of active volcanoes.

7 The formation of rocks, minerals, and geologic structures

Volcanoes create environments. This was demonstrated very dramatically, and shown on television, in 1963, when a new submarine volcano called Surtsey (volcano.und.nodak.edu/vwdocs/volc_images/europe_west_asia/surtsey.html) erupted to the south of Iceland. The eruption was extremely violent, because sea water entered the open volcanic vent, and steam, gas, pieces of rock, and ash were hurled many kilometres into the air. Since then eruptions of this type have been called ‘Surtseyan’. The lava cone was high enough to rise above the surface, where it formed what is now the island of

Surtsey. As it cooled, sea birds began to settle on it.³ They carried plant seeds and slowly plants and animals began to colonize the new land.

Even the damage caused by destructive eruptions is repaired, although this can take a long time. The 1883 eruption of Krakatau, in the Sunda Strait between Java and Sumatra, Indonesia, destroyed almost every living thing on Krakatau itself and on two adjacent islands. Three years later the lava was covered in places by a thin layer of cyanobacteria, and a few mosses, ferns, and about 15 species of flowering plants, including four grasses, had established themselves. By 1906 there was some woodland, which is now thick forest. The only animal found in 1884 was a spider, but by 1889 there were many arthropods and some lizards. In 1908, 202 species of animals were living on Krakatau and 29 on one of the islands nearby, although bats were the only mammals. Rats were apparently introduced in 1918. Species continued to arrive and 1100 were recorded in 1933 (KENDEIGH, 1974, pp. 24–25).

Rock that forms from the cooling and crystallization of molten magma is called ‘igneous’, from the Latin *igneus*, ‘of fire’, and all rock is either igneous or derived from igneous rock. This must be so, since the molten material in the mantle is the only source for entirely new surface rock. If the magma reached the surface before cooling the rock is known as ‘extrusive’; if it cooled beneath the surface surrounded by older rock into which it had been forced, it is said to be ‘intrusive’. Intrusive rock may be exposed later as a result of weathering. It is not only igneous rocks that can form intrusions. Rock salt (NaCl) can accumulate in large amounts beneath much denser rocks and rise through them very slowly to form a salt dome. Salt domes are deliberately sought by geologists prospecting for oil but occasionally they can break through the surface. When this happens the salt may flow downhill like a glacier.

The character of the rock depends first on its chemical composition. If it is rich in compounds of iron and magnesium it will be dark (melanocratic); if it is rich in silica, as quartz and feldspars, it will be light in colour (leucocratic). Rock between the two extremes is called ‘mesocratic’. The rock comprises minerals, each with a particular chemical composition, and minerals crystallize as they cool. Whole rock is quarried for building and other uses; many minerals are mined for the chemical substances they contain, especially metals, and some are valued as gemstones. Crystallization proceeds as atoms bond to particular sites on the surface of a seed crystal, forming a three-dimensional lattice. It can occur only where atoms have freedom to move and so the more slowly a molten rock cools the larger the crystals it is likely to contain. The crystal size gives the rock a grain structure, which also contributes to its overall character. The type of rock is also determined by the circumstances of its formation. Lava that flows as sheets across the land surface or sea bed often forms basalt, a dark, fine-grained, hard rock. Basalt covers about 70 per cent of the Earth’s upper crust, making it the commonest of all rocks; most of the ocean floor is of basalt overlain by sediments and on land it produces vast plateaux, such as the Deccan Traps in India. Intrusive igneous rocks are usually of the light-coloured granite type. Beyond this, however, the identification and classification of igneous rocks are rather complicated.⁴

Rocks formed on the ocean floor may be thrust upward to become dry land or exposed when the sea level falls. Tectonic plate movements are now believed to be the principal mechanism by which this occurs. Where two plates collide the crumpling of rocks can raise a mountain chain, as is happening now between the Indian and Eurasian plates, raising the Himalayan chain. The Himalayas, which began to form some 52–49 million years ago following the closure of the Tethys Sea, are linked to the Alps, which began forming about 200 million years ago owing to very complex movements of a number of plates (WINDLEY, 1984, pp. 202–308). The formation of a mountain chain by the compression of crustal rocks is known as an ‘orogeny’ (or ‘orogenesis’).

The British landscape was formed by a series of orogenies. The first, at a time when Scotland was still joined to North America, began about 500 million years ago and produced the Caledonian-Appalachian mountain chain (WINDLEY, 1984, pp. 181–208) as well as the mountains of northern Norway. The Appalachians were later affected by the Acadian orogeny, about 360 million years ago, and the Alleghanian orogeny, about 290 million years ago. Europe was affected by the Hercynian and Uralian orogenies, both of which occurred at about the same time as the Alleghanian. Figure 2.3 shows the area of Europe affected by several orogenies.⁵

Igneous intrusions can be exposed through the weathering away of softer rocks surrounding them. Such an exposed intrusion, roughly circular in shape and with approximately vertical sides, is called a ‘boss’ if its surface area is less than 25 km² and a ‘batholith’ if it is larger (and they are often much larger). Dartmoor and Bodmin Moor, in Devon and Cornwall, Britain, lie on the surface of granite batholiths.

Mountains are not always formed from igneous rocks, however. There are fossil shells of marine organisms at high altitudes in the Alps and Himalayas, showing that these mountains were formed by the crumpling of rocks which had formed from sea-bed sediments.

Many sedimentary rocks are composed of mineral grains eroded from igneous or other rocks and transported by wind or more commonly water to a place where they settle. Others, said to be of ‘biogenic’ origin, are derived from the insoluble remains of once-living organisms. Limestones, for example, are widely distributed. Most sediments settle in layers on the sea bed, to which rivers have carried them. Periodic changes in the environmental conditions in which they are deposited may cause sedimentation to cease and then resume later, and chemical changes in the water or the sediment itself will be recorded in the sediments themselves and in the rocks into which they may be converted.



Figure 2.3 *The mountain-forming events in Europe*

Note: The thick lines (- • - • -) mark the Alpine orogeny

Sandstones are perhaps the most familiar sedimentary rocks, consisting mainly of sand grains, made from quartz (silica, SiO_2) which crystallized originally into igneous rock. Clay particles, much smaller than sand grains, can pack together to make mudstones. Sediments rich in calcium carbonate, often consisting mainly of the remains of shells and containing many fossils, form limestone and dolomite (sometimes known as ‘dolostone’ to distinguish it from the mineral called dolomite) (HOLMES, 1965, ch. VI, pp. 118–141). Particles deposited as sediments are changed into rock by the pressure of later deposits lying above them and the action of cementing compounds subsequently introduced into them. The process, occurring at low temperature, is called ‘diagenesis’. Some sedimentary rocks are very hard and many, especially sandstones and limestones, make excellent and durable building stone. Once formed, a sedimentary rock is subject to renewed weathering, especially if it is exposed at the surface, so sedimentary rocks continually form and re-form.

Sediments are deposited in horizontal layers, called ‘beds’, but subsequent movements of the crust often fold or fracture them. It is not unusual for beds to be folded until they are upside down, and the reconstruction of the environmental conditions under which sediments were deposited from the study of rock strata often begins by seeking to determine which way up they were when they formed. All in all, the interpretation of sedimentary structures can be difficult.⁶ Figure 2.4 shows the sequence of events by which sedimentary structures may be folded, sculptured, and then subside to be buried beneath later beds producing an unconformity.

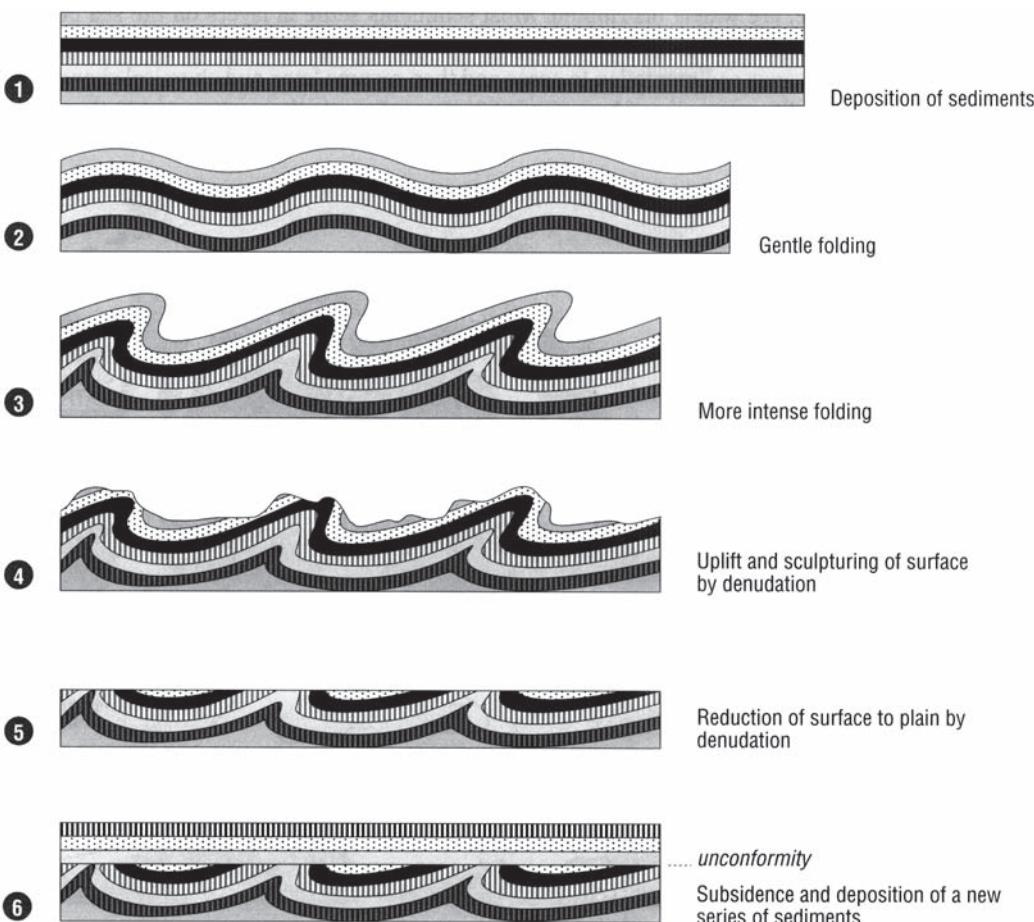


Figure 2.4 Stages in the development of an unconformity

The extreme conditions produced by the folding and shearing of rock can alter its basic structure by causing some of its minerals to recrystallize, sometimes in new ways. This process, called ‘metamorphism’, also happens when rock of any type comes into contact with molten rock, during the intrusion of magmatic material for example. Marble is limestone or dolomite (dolostone) that has been subjected to metamorphism at high temperature. Such shells as it contained are completely destroyed as the calcium carbonate recrystallizes as the mineral calcite. If quartz or clay particles are present, new minerals may form, such as garnet and serpentine. Hard limestone containing fossils is often called marble, but there are no fossils in true marble.

Slate is also a metamorphic rock, derived from mudstone or shale, in which the parallel alignment of the grains, due to the way the rock formed, allows the rock to cleave along flat planes (HOLMES, 1965, pp. 168–170). It may contain fossils, although they are uncommon and usually greatly deformed, because slate forms when the parent sedimentary rock is squeezed tightly between two bodies of harder rock that are moving in parallel but opposite directions, so its particles, and fossils, are dragged out. It is this that gives slate its property of ‘slaty cleavage’ which, with the impermeable surface imparted at the same time, makes it an ideal roofing and weatherproofing material. Metamorphic rocks are widely distributed and with practice you can learn to recognize at least some of them.⁷

All the landscapes we see about us and the mineral grains that are the starting material for the soils which form over their surfaces are produced by these processes. The intrusion or extrusion of igneous rock supplies raw material. This weathers to provide the mineral grains which become soil when they are mixed with organic matter, or is transported to a place where it is deposited as sediment. Pressure converts sediments into sedimentary rocks, which may then be exposed by crustal movements, so that erosion can recommence. Metamorphic rocks, produced when other rocks are subjected to high pressures and/or temperatures, are similarly subject to weathering. It is the cycling of rocks, from the mantle and eventually back to it through subduction, that produces the physical and chemical substrate from which living organisms can find subsistence.

8 Weathering

No sooner has a rock formed than it becomes vulnerable to attack by weathering. The word ‘weathering’ is slightly misleading. We associate it with wind, water, freezing, and thawing. These are important agents of weathering, but they are not the only ones. Weathering can be chemical as well as physical and it often begins below ground, completely isolated from the weather.

Beneath the surface, natural pores and fissures in rocks are penetrated by air, containing oxygen and carbon dioxide, and by water into which a wide variety of compounds have dissolved to make an acid solution. Depending on their chemical composition, rock minerals may dissolve or be affected by oxidation, hydration, or hydrolysis (HOLMES, 1965, pp. 393–400). Oxidation is a reaction in which atoms bond with oxygen or lose electrons (and other atoms gain them, and are said to be ‘reduced’). Hydration is the bonding of water to another molecule to produce a hydrated compound; for example, the mineral gypsum ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) results from the hydration of anhydrite (CaSO_4). Hydrolysis (lysis, from the Greek *lusis*, ‘loosening’) is a reaction in which some parts of a molecule react with hydrogen ions and other parts with hydroxyl (OH^-) ions, both derived from water, and this splits the molecule into two or more parts.

The result of chemical weathering can be seen in the limestone pavements found in several parts of England, Wales, and Ireland.⁸ South Devon, England, is famous for its red sandstones, well exposed in the coastal cliffs of the Torbay area. These date from the Devonian Period, some 400 million years

ago, when what is now Devon was a hot, arid desert. The desert sand contained some iron, which was oxidized to its insoluble red oxide, giving the sandstone its present colour.

Limestone pavement

A distinctive feature, sometimes covering a large area, that occurs in many parts of the world. It forms when horizontal limestone beds are exposed by the erosion of any material that may once have covered them and joints within them are penetrated by rain water carrying dissolved CO₂. This weak carbonic acid (CO₂ + H₂O → H⁺ + (HCO₃)⁻) reacts with calcium carbonate to produce calcium bicarbonate, which is soluble in water and is carried away. This widens the joints to form deep crevices (called 'grikes' in England) separated by raised 'clints'. Small amounts of soil accumulating in the sheltered grikes provide a habitat for lime-loving plants, making limestone pavements valuable botanically. At a deeper level, the grikes may join to form caves. Particular areas of limestone pavement are protected in Britain by Limestone Pavement Orders issued under the Wildlife and Countryside Act 1981, mainly to prevent the stone being taken to build garden rockeries and for other ornamental uses.

Iron oxidizes readily and this form of weathering has produced hematite (Fe₂O₃), one of the most important iron ore minerals, some of which occurs in banded ironstone formations, 2–3 billion years old, composed of alternating bands of hematite and chert (SiO₂). Iron and other metals can also be concentrated by hydrothermal, or metasomatic, processes. Near mid-ocean ridges, where new basalt is being erupted on to the sea bed, iron, manganese, and some other metals tend to separate from the molten rock and are then oxidized and precipitated, where particles grow to form nodules, sometimes called 'manganese nodules' because this is often the most abundant metal in them. Vast fields of nodules, containing zinc, lead, copper, nickel, cobalt, silver, gold, and other metals as well as manganese and iron, have been found on the floor of all the oceans (KEMPE, 1981). A few years ago serious consideration was given to the possibility of dredging for them, but at present metals can be obtained more cheaply by conventional mining on land.

Hydrothermal weathering, in which hot solutions rise from beneath and react with the rocks they encounter, produces a range of commercially valuable minerals, perhaps the best known of which is kaolin, or 'china clay'. This material was first discovered in China in 500 BC and was used to make fine porcelain, hence the names 'china clay' and 'kaolin', from *kao ling*, meaning 'high ridge', the type of landscape in which it occurred. Today it is still used in white ceramics, but most is used as a filler and whitener, especially in paper. The paper in this book contains it. Kaolin deposits (www.wbb.co.uk/) occur in several countries, but the most extensively mined ones are in Cornwall and Devon, Britain.

Kaolin is a hydrated aluminium silicate, Al₂O₃2SiO₂·2H₂O, obtained from the mineral kaolinite. The British deposits occur in association with the granite batholiths and bosses intruded during the Hercynian orogeny. Granites consist of quartz crystals, mica, and feldspars. Feldspars are variable in composition. All are aluminium silicates, those associated with the kaolinite deposits being plagioclase feldspars, relatively rich in sodium. As the intruded granite was cooling, it was successively exposed to steam, boron, fluorine, and vaporized tin. The feldspar reacted with these, converting it into kaolinite (the process is known as kaolinization), a substance consisting of minute white hexagonal plates

which are separated from the rock industrially by washing and precipitation, leaving a residue of quartz grains (a white sand) and mica. About 15 per cent of the material is recovered as kaolin (www.kaolin.com/ccpmin.htm), 10 per cent is mica waste, and 75 per cent is sand, which is also waste although it has found some use for building and landscaping. In some places the kaolinization process has been completed from above, possibly by humic or other acids from overlying organic material, but most of the kaolinite formed at depth is overlain by unaffected granite, probably because the upward movement of acidic fluids was halted by the absence of veins or joints it could attack. The resulting deposits are funnel-shaped, extending in places to depths of more than 300 m.

Bauxite, the most important ore of aluminium, is also produced by the chemical weathering of feldspars, in this case by hydration. Bauxite is a mixture of hydrous aluminium oxides and hydroxides with various metals as impurities; to be suitable for mining it should contain 25–30 per cent of aluminium oxide.

Bauxite is a variety of laterite, one product of the kind of extreme weathering of soil called ‘laterization’. The word ‘laterite’ is from the Latin *later*, meaning ‘brick’, and laterite is brick-hard. Laterization occurs only in some parts of the seasonal tropics, where soils are derived from granite parent material, but it is possible that removing the forest or other natural vegetation in such areas may trigger the formation of laterites. These can be broken by ploughing.

Except on steep slopes, tropical soils overlying granite can be up to 30 m deep. Naturally acidic water from the surface percolates through them, steadily eating away at the parent rock beneath, and plants draw the water up again through their roots. Water is also drawn upward by capillary attraction through tiny spaces between soil particles and evaporates from the surface. If the rainfall is fairly constant through the year, the movement of water is also constant, but if it is strongly seasonal, evaporation exceeds precipitation during the dry season and mineral compounds dissolved in the soil water are precipitated, the least soluble being precipitated first. Provided vegetation cover is adequate, with roots penetrating deep into the soil, the minerals will not accumulate in particular places and when the rains return they will be washed away. If there is little plant cover, however, they may accumulate near the surface. The most insoluble minerals are hydroxides of iron and aluminium (kaolinite) and they are what give many tropical soils their typically red or yellow colour (HOLMES, 1965, pp. 400–401). Soil developed over granite will contain sand, or quartz grains, and clays derived from feldspars in varying amounts. Figure 2.5 shows how these can grade almost imperceptibly from one to the other and from both into laterite. Laterite layers or nodules are hard, but not usually thick, because, being impermeable, they prevent further percolation of water downwards into the soil and thus bring the laterization process to an end. Erosion of the surface layer may then expose the laterite.

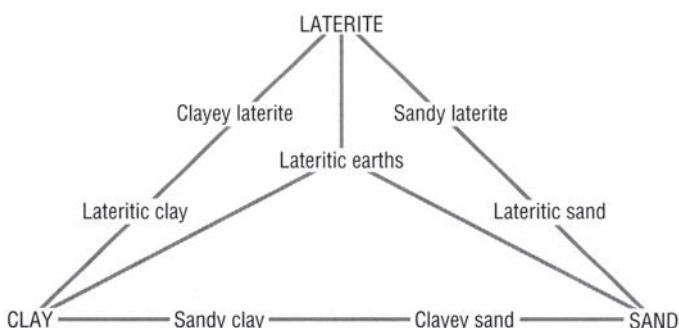


Figure 2.5 Gradation of clay and sand to laterite

Source: Holmes, Arthur. 1965. *Principles of Physical Geology*. Nelson, Walton-on-Thames

Laterization does not necessarily render a soil useless and many relatively laterized soils are cultivated, although some soils resembling lateritic soils, for example in parts of the eastern United States, are not truly laterized. Indeed, there are doubts about the extent to which laterization is occurring at present. Lateritic soils in the West Indies, Indonesia, Australia, India, and China may well be of ancient origin (HUNT, 1972, p. 193).

Living organisms contribute to weathering. By moving through soil they assist the penetration of air and water, and the decomposition of organic material releases acids and carbon dioxide, some of which dissolves into the soil water. Biological activity contributes greatly to the formation of soil.

Physical weathering is also important in soil formation, especially in its initial stages, but it can also degrade soils through erosion. Thermal weathering, which is the expansion and contraction due to repeated heating and cooling, causes rocks to flake, especially if water is held within small crevices. Small particles detached from the rock may then be carried by the wind and if they strike other rocks more particles may be chipped from them. Depending on their size, the particles may be carried well clear of the ground or may roll and bounce along the surface; the process is called ‘saltation’. Most serious erosion is due to water, however. All water flowing across the land surface carries soil particles with it. This can lead to the formation of rills and gullies into which more particles are washed and then transported, or where water flows as sheets whole surface layers can be removed. In addition to this, all rivers erode their banks, and waves erode the shores of lakes and the sea (HUDSON, 1971, pp. 33–46).

These processes are entirely natural and part of the cycles by which originally igneous rocks are converted into sediments and landforms are made and age, but human activities can accelerate them. The UN estimates that in the world as a whole, some 1.093 billion hectares (ha) of land have been degraded by water erosion, 920 million ha by sheet and surface erosion and a further 173 million ha by the development of rills and gullies. Of the total area subject to serious degradation by water, 43 per cent is attributed to the removal of natural vegetation and deforestation, 29 per cent to overgrazing, 24 per cent to poor farming practices, such as the use of machinery that is too heavy for the soil structure to support and the cultivation of steep slopes, and 4 per cent to the over-exploitation of vegetation (TOLBA AND EL-KHODY, 1992, pp. 149–150). There is, however, some evidence that modern farming techniques can reduce soil erosion substantially. A study of a site in Wisconsin found that erosion in the period 1975–93 was only 6 per cent of the rate in the 1930s (TRIMBLE, 1999). This may be due to higher yields from the best land, combined with methods of tillage designed to minimize erosion (AVERY, 1995).

Weathering is the general name given to a variety of natural processes by which rock is recycled and soil and landscapes created. It creates and alters environments, but human activities can accelerate it on vulnerable land, degrading natural habitats and reducing agricultural productivity.

9 The evolution of landforms

The weathering of exposed rocks and the erosion and transport of loose particles create the landscapes we see and change them constantly. Change is usually slow, but not always. The 1952 Lynmouth flood was very sudden (see box), but not far away there are landscapes which record conditions long ago. During the most recent glaciation the ice sheets did not extend as far south as Devon, but on the high granite batholith of Dartmoor the climate was severe, with permanently frozen ground (permafrost), and to this day parts of Dartmoor are periglacial landscapes. Rock masses were shattered by the repeated freezing and thawing of water that penetrated crevices.

In winter the water expanded as it froze, widening the crevices, and in summer the water shrank as it melted, releasing flakes of rock and also large boulders. For those few weeks in summer when the weather was warm enough to thaw the surface layers of the permafrost, turning soil locked solid by ice into wet mud, the mud, together with large boulders embedded in it, slid downhill, only to be brought to a halt when the temperature dropped and the mud froze again. Today, although there is no permafrost, the scattering of boulders around the tors remains as a record of the climate more than 10000 years ago. Similar periglacial processes acting on the weak, jointed chalk of southern England caused slopes to retreat through the loss of material from their faces and produced large deposits of the angular debris comprising fragments of varying sizes called ‘coombe rock’ or sometimes ‘head’ (other definitions confine ‘head’ to deposits other than chalk). There are similar periglacial relics in North America and elsewhere in Europe.

The Lynmouth flood

On Exmoor, in south-west England, in the summer of 1952, almost 230 mm of rain fell in 24 hours on to land that was already waterlogged. The water drained northward, carried in two rivers, the East and West Lyn, which enter the sea together at the small village of Lynmouth, falling some 300 m in rather more than 1 km. Unable to carry the volume of water, during the night of 15 and 16 August both rivers flooded and the overflow from the West Lyn cut a new channel that took it through Lynmouth, rejoining the original course at the mouth. Houses, roads, and bridges were destroyed, an estimated 40000 tons of trees, soil, boulders, and rubble and masonry from collapsed structures piled up in the village, and 31 people were killed. The disaster was caused by nothing more than rain. Lynmouth was subsequently restored and is now a popular and attractive holiday resort.

Present permafrost regions occur in latitudes much higher than Britain. In Canada and Alaska within the Arctic Circle in places the permafrost is 400 m thick and in parts of Siberia it is 700 m thick. In Resolute Bay, in the Canadian Arctic, it extends to a depth of about 1000 m. Overall, nearly 20 per cent of the land area within the Arctic Circle is permafrost, and has remained in this state since the retreat of the ice sheets that once covered it.

Ice sheets are major sculptors of landscapes. As they move, they scour away all soil and other loose material, pushing it ahead and to the sides of them, where it may form moraines. They smooth angular rocks and the weight of the ice depresses the ground beneath. During a major glaciation ice sheets may grow to a thickness of more than 2500 m and depress the underlying surface by 600 m, which may take it to below sea level. As the ice retreats, the surface rises again, but it is a slow process, at least as measured on a human scale. Northern Canada, where shore-lines rose several tens of metres in less than 1000 years, and Scandinavia are still rising to compensate for the loss of their ice sheets around 10000 years ago; in Scandinavia the surface was depressed by about 1000 m and has subsequently risen by 520 m. This ‘glacioisostasy’ demonstrates the slight flexibility of the Earth’s crust.

Because there is a lag between the disappearance of the ice sheets and the recovery of the original surface elevation, bowls may remain where the ice was thickest. Depending on their location, these may be flooded by the sea or fill with fresh water. The North American Great Lakes and the Baltic Sea were made in this way. On a much smaller scale, so were the lakes of the English Lake District.

Ice accumulating in a pre-existing hollow will erode the sides to the open-sided, approximately circular shape of a cirque (also known as a ‘corrie’ or ‘cwm’). Where a relatively narrow glacier flows into the sea the trough it excavates may later form a fjord, known in Scotland as a ‘sea loch’. Some fjords are more than 1200 m deep. In latitudes higher than about 50°, ice has been the major geomorphological (‘landscape-forming’) agency.

Soil will tend to move slowly downslope by ‘soil creep’, caused by the expansion and contraction of material due to repeated wetting and drying, or ‘solifluction’, where the soil is lubricated by rain water (formerly the term ‘solifluction’ was applied only to periglacial environments where the ground is frozen for part of the year, but it is now used more widely and is recognized as an important process in some tropical areas). The rate of soil creep has been measured in the English Pennines as between 0.5 and 2.0 mm at the surface and 0.25 to 1.0 mm in the uppermost 10 cm (SMALL, 1970, p. 224). If the soil is deep and the underlying rock extensively weathered, large masses may slip suddenly and move rapidly as ‘earth flows’. The collapse of coal tips at the Welsh village of Aberfan in 1966 was of this type (in this case known strictly as a ‘flowslide’). The tips had been built over springs. Tip material absorbed the water, greatly increasing its weight but simultaneously lubricating it until it lost its inertia catastrophically (SMALL, 1970, p. 29–34). Earthquakes can break the bonds holding soil particles together, resulting in earth flows of dry material.

There are several ways in which masses of rock and earth can move downslope (HOLMES, 1965, p. 481). All such movements alter the shape of slopes, generally smoothing and reducing them. Figure 2.6 shows the stages by which this happens:

(1) material from the free face is detached and falls to form a scree which buries a convex lower slope; (2) further falls cause the free face to retreat until it disappears altogether, leaving a slope that grades smoothly to the level of the higher ground; (3) the slope itself then erodes further. It can also happen that accumulated water increases the weight of a mass of weathered material until it shears away and slides down a concavely curved shear plane between it and the adjacent material. Because of the curved slope, the sliding layers are tilted backwards as they descend, so the toe of the slide is tipped upwards, forming a barrier behind which further debris will be held. This is a ‘rotational slide’, examples of which can be seen in several places along the south coast of England and in the Isle of Wight. Most failures are quite complex and involve more than one mechanism.

Geomorphology, the study of landforms and the processes by which they are produced and change, began with the work of an American geologist, William Morris Davis (1850–1934), of Harvard University. He proposed that landscapes evolve through a ‘cycle of erosion’ (the ‘Davisian cycle’). This begins

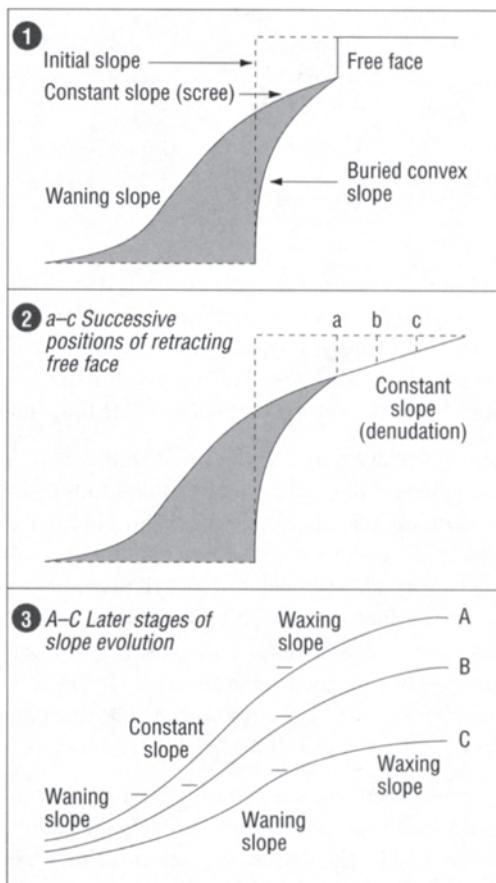


Figure 2.6 **Slope development**

when land is raised by tectonic movements. In a young landscape hills slope steeply and the slope of river beds is irregular. As the landscape matures, hill slopes become gentle and river beds slope smoothly. Eventually, the old landscape has eroded to a gently rolling peneplain ('almost a plain'), a word Davis coined. His idea was (posthumously) challenged, in 1924, by the German geologist Walther Penck (1888–1923), who argued that once a slope has settled at an angle which is mechanically stable for the material of which it is composed, it will maintain that angle. Erosion will wear away its face, but will not make it more shallow, so the face will retreat but the angle will remain fairly constant, and the steeper the slope the faster it will erode, because the slow-moving weathered material on a shallow slope will protect the underlying surface. Thus, if a slope is steeper near the bottom than it is higher up, the lower slope will erode faster than the upper slope and the structure will collapse. As the argument developed, geomorphologists came to realize that a true understanding of 'the slope problem' can best be gained from studies of low-latitude landscapes that have not been formed mainly by glacial action, as were those on which the theories of Davis and Penck were largely based (SMALL, 1970, pp. 194–224). Interest in the topic is not purely academic, for an understanding of how rock and soil behaves on sloping ground is necessary for engineers calculating the risks of landslides, erosion, and flooding, and devising schemes to minimize them. The matter is of major environmental importance.

Rivers provide the principal means by which particles eroded from surface rocks are transported from the uplands to the lowlands and eventually to the sea. Rivers are also major landscape features in their own right and, by cutting channels across the surface, important agents in the evolution of landscapes. It is not only mineral particles they transport, of course. Water draining into a river from adjacent land

also contains organic matter and dissolved plant nutrients, and rivers also carry those substances we discharge into them as an apparently convenient method of waste disposal. They are also a major source of water supplied for domestic and industrial use.

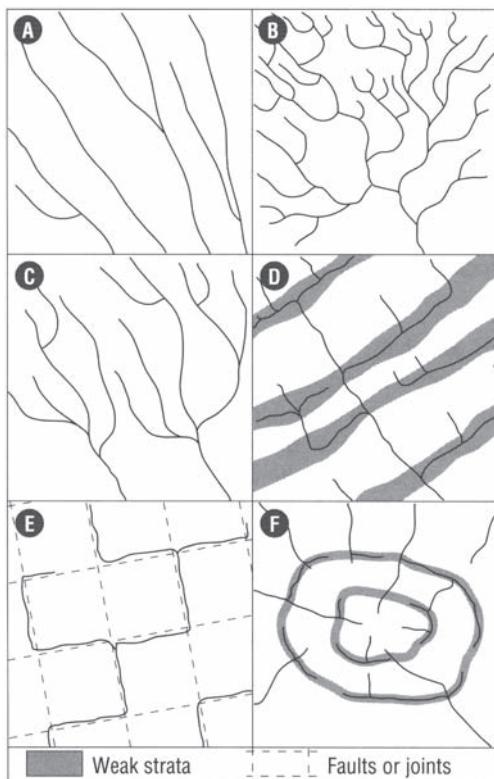


Figure 2.7 Drainage patterns. A, subparallel; B, dendritic; C, semi-dendritic; D, trellised; E, rectangular; F, radial

Water drains from higher to lower ground, moving slowly as ground water between the freely draining soil and an impermeable layer of rock or clay, eventually emerging at the surface as a spring, seeping from the ground, or feeding directly into a river. The 'water table' is the upper limit of the ground water, below which the soil is fully saturated. These terms have the same meanings in British and North American usage, but confusion can arise over 'watershed', which has two different meanings. A drainage system removes water from a particular area, and one such area is separated from adjacent areas. In Britain, the area from which water is removed by a particular drainage system is called a 'catchment' and in North America it is called a 'watershed'. One catchment is separated from another by a 'divide', which in Britain is sometimes known as a 'watershed', and within a catchment the drainage system forms a pattern. Figure 2.7

illustrates six of the commonest patterns, but others are possible and real patterns are seldom so clearly defined as the pictures may suggest. Climate, the type of rock, and the extent of erosion all play a part in determining the type of pattern that will develop. Dendritic patterns, for example, usually form on gently sloping land of fairly uniform geologic character. Radial patterns occur around domed hills and batholiths, and trellis patterns where rivers cross, more or less at right angles, alternating bands of relatively hard and soft rocks.

As they flow, rivers can be approximately classified into zones, mainly on a biological basis. The headstream, or highland brook, is small, usually torrential (which means its water flows at more than 90 cms⁻¹), and the water temperature varies widely. Few aquatic animals can survive in it. A little lower, trout can survive in what is still a fast-flowing stream, the troutbeck. Silt and mud begin to collect at the bottom of the minnow reach or grayling zone, some plants can survive, and the animal life becomes a little more diverse. In the lowland reach or bream zone the water flows slowly, the river is often meandering, and animal life is diverse. In this final zone the river flows across the coastal plain into the estuary.

10 Coasts, estuaries, sea levels

It seems natural to think of an estuary in terms of the river flowing into it, to see it as the end of the river, with a boundary somewhere offshore where the river meets and merges with the sea. Stand on a headland overlooking an estuary and this is how it looks, but the picture is misleading. An estuary is more accurately described as an arm of the sea that extends inland and into which a river flows. An estuary is dominated by the sea rather than its river, and many estuaries are in fact 'rias', or 'drowned river valleys', old river valleys which were flooded at some time in the past when the sea level rose. The estuaries of south-west England are good examples of rias. In several cases, such as the Camel in north Cornwall, before the marine transgression that began about 10300 years ago the sea was 36 m below its present level (the sea is still rising at about 25 cm per century), and gently undulating land, with hills formed from igneous intrusions through Devonian slate which survive now as offshore islands, extended up to 5 km from the present coast. This land was blanketed with mixed deciduous forest. Remnants of the forest have been found on the sea bed at several points along the coast and its botanical and faunal composition determined (JOHNSON AND DAVID, 1982).

Sea levels change and at various times in the past they have been both higher and lower than they are today, and they are changing still. During glacial periods (ice ages), sea levels fall, because the volume of the oceans decreases as water evaporated from them accumulates in ice sheets. As the weight of ice depresses the land beneath it sea levels rise; as the ice sheets melt they also rise; and as land depressed by the weight of ice rises again when the ice has melted they fall. There is clear evidence in many places that sea levels were much lower at some time in the past. Raised beaches can be found that are several metres above the present high-tide level. These are areas of approximately level ground, nowadays usually vegetated, containing large numbers of shells of marine organisms. They can have been produced only by the movement of waves and tides over them, at a time when they formed the shore; they are ancient beaches now some distance from the sea.

The sea bed at the mouth of the Camel estuary is mainly sandy, with sand bars, and there are many sandy beaches along the adjacent coast. Sand consists primarily of quartz grains weathered and eroded from igneous rocks inland and transported by the river. They are deposited at the mouth of the estuary, then transported further by tides and sea currents. As they move they become mixed with varying amounts of sea shells, most of which are crushed to tiny fragments through being battered by harder

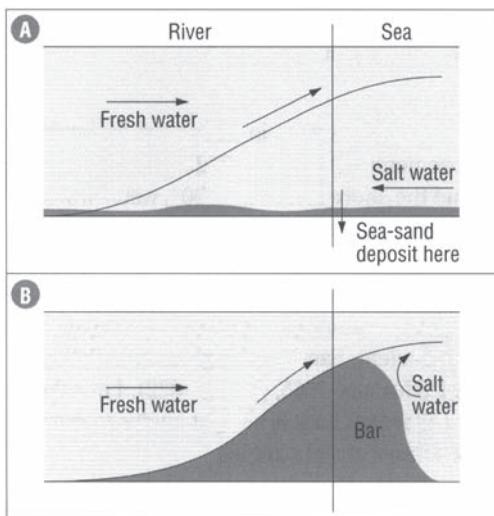


Figure 2.8 Deposition of sand and formation of an estuarine sand bar

As the fresh water is forced to rise, it loses energy, and the quantity of material a river can transport, known as its ‘traction load’ or ‘bed load’, is directly proportional to the energy with which it flows. This depends in turn on such factors as the potential gravitational energy causing the water to flow (essentially the height of the river source above sea level), the gradient of the channel, and the amount of friction caused by contact with the banks and bed (SMALL, 1970, pp. 34–41). As the river water rises and loses energy, the sand grains sink, falling through the underlying salt water and on to the bed. Figure 2.8 shows what happens and how over time the deposition of sand can lead to the formation of a bar at the estuary mouth.

Sand does not remain static on the sea bed and material for the growth of a bar is also provided by sand being carried landward or along the shore by tides and sea currents. The sea water also loses energy as it pushes against the fresh water, and the sand it carries is again deposited.

Sand grains are much larger and heavier than the particles of silt rivers also carry. Silt particles are 2–60 micrometres (μm) in diameter, sand grains 60–2000 μm (in the British standard classification; in the widely used Udden-Wentworth classification they are 4–62.5 μm and 62.5–2000 μm respectively). Ordinarily, large particles would be expected to settle first and small ones later, but in an estuary the opposite occurs. Mudbanks, composed of silt, still smaller clay particles and, mixed with them, organic molecules from the decomposition of the waste products and dead bodies of biological organisms, form inland of the sand banks. Flocculation is the process responsible for this phenomenon. Many of the very small particles carry an electrical charge owing to the presence of bicarbonate (HCO_3^-), calcium (Ca^{2+}), sulphate (SO_4^{2-}), and chlorine (Cl^-) ions. In the boundary zone where fresh and salt water meet, these particles encounter chlorine, sodium (Na^+), sulphate, and magnesium (Mg^{2+}) ions, which bond to them and attract more silt particles, so the material forms clumps larger and heavier than sand grains, and these settle. The organic material mixed with them provides rich sustenance for bacteria and, closer to the surface, burrowing invertebrate animals, which provide food for wading birds. The environment is harsh because the salinity of the water varies widely, so although only a restricted number of species can regulate their osmosis well enough to survive in the mud, those which succeed do so in vast numbers. Estuarine waters may also be enriched by a ‘nutrient trap’, where the current pattern causes dissolved plant nutrients to be retained

stones, producing a beach material with a relatively high calcium carbonate content; this was formerly used by farmers as ‘lime’, to raise the pH of their soils.

Sand that has been transported many miles by river is deposited where fresh water and sea water meet. Because sea water is denser than fresh water the two do not mix readily and tend to flow in separate channels. The configuration of these channels is determined by the topography of the estuary itself; they may flow side by side or form a wedge, in which fresh water rises over the sea water. On an incoming tide, freshwater and seawater currents often flow in opposite directions and marine fish can move considerable distances inland by keeping to the salt-water channel.

(CLARK, 1977, p. 6). In sheltered areas, plants rooted in the mud trap further sediment. In this way mangroves extend some tropical coastlines seaward and, in temperate regions, sediment trapped by saltmarsh vegetation raises the surface until it is beyond the reach of the tides and becomes dry land.

Beach material is transported by currents produced mainly by waves in directions determined by the angle between the shoreline and prevailing movements of the water. Tides have much less effect. If the angle between the approaching waves and the shore is less than 90°, water will flow parallel to the shore, carrying loose beach material with it. This is longshore drift. Its mechanics are complex, but its effect is to shift material to one end of the beach, where it may be trapped against resistant material, such as a cliff or groynes built to prevent beach erosion, swept away entirely (and commonly deposited elsewhere), or deposited where the shoreline turns into the mouth of a bay or estuary and the current becomes turbulent. Sand or shingle accumulating in this way may eventually rise above the water as a spit (HOLMES, 1965, pp. 820–822). Longshore drift moves material in such a way as to rearrange the shoreline so it is at right angles to prevailing waves. This is why waves often appear to meet the shore at right angles.

Waves are produced mainly by wind. This is obvious enough in the case of the huge waves thrown against the coast during storms, but less so in the case of the gentle swell that moves against ocean shores on even the calmest days and in fact is caused by strong winds 1500 km or more from land. The size of sea waves depends on wind force and the length of time during which it blows, the distance over which the water is affected (called the ‘fetch’), and the influence of waves that were generated elsewhere. Some waves are also driven gravitationally. Low atmospheric pressure allows the sea level to rise beneath it and water flows downhill towards areas where higher pressure produces a lower sea level.

Beaches are built by waves, and especially by ‘spilling breakers’ which move a long way up the beach before breaking, spilling a little water from their crests. ‘Plunging breakers’, which collapse earlier, produce a strong backwash that tends to erode the beach. Where the usual wave pattern is of spilling breakers arriving at about six to eight a minute, the uppermost part of the beach is likely to end in a raised mound of coarse sand or pebbles (a ‘berm’).

Where waves crash against a vertical rock face, their force is considerable. It has been measured at up to 25 t m⁻² (SMALL, 1970, p. 438). If the rock is soft, or has many joints or fissures, this is enough to erode it. In harder rock, it is enough to violently compress air held in crevices and allow the air to expand again as the water recedes. This weakens the rock and may produce further cracks into which air enters, until eventually sections are detached as boulders.

Sea cliffs are the result of the wave erosion of hills and as what used to be a hill is cut back, the base becomes a gently sloping wave-cut platform and the eroded material accumulates just below the low-tide limit as a wave-built terrace. Figure 2.9 illustrates the process and shows that it is the ultimate fate of sea cliffs to be eroded completely, until the land slopes gently from the upper limit of wave action to the low-tide line.

How long this takes depends on the resistance offered by the rock, the degree to which it is exposed to the full force of the waves, and the topography of the original high ground. In north Cornwall, Britain, the very impressive sea cliffs have taken around 10000 years to reach their present condition.

The process may never be completed, because the sea level may reverse its present trend and fall again. This could happen were the ice sheets to advance in a new glaciation. Alternatively, cliff erosion may accelerate as the sea level rises. In some places the present sea-level rise is due to erosion, so it is really a sinking of the land rather than a rising of the sea. It is also believed to be due, in some places exclusively, to the expansion of sea water due to a warming of the sea as

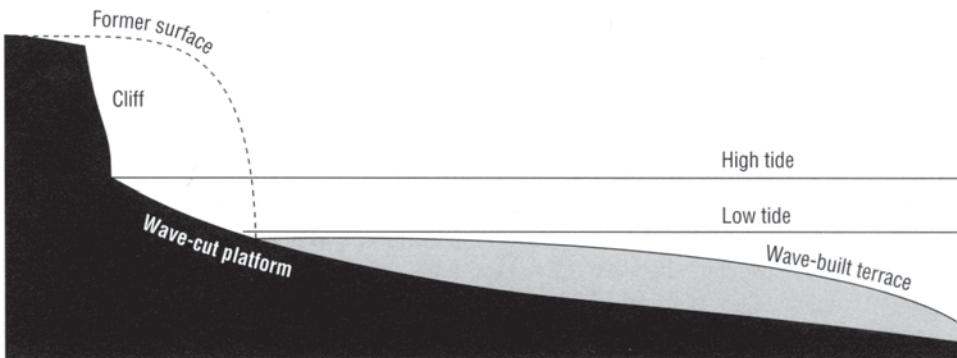


Figure 2.9 The development of a sea cliff, wave-cut platform, and wave-build terrace

a result of what may be a general climatic warming. Many climatologists predict that such a warming is likely to continue, but the consequences for sea levels are difficult to estimate and predictions vary widely.

11 Energy from the Sun

Tides are driven by gravitational energy and plate tectonics by the heat generated by the radioactive decay of elements in the Earth's mantle, but the energy driving the atmosphere, oceans, and living organisms is supplied by the Sun. To a limited extent this energy can also be harnessed directly to perform useful work for humans. Solar heat can be used directly to warm buildings and water, desalinate water, and cook food. Sunlight can be converted into electricity. Electrical power can also be generated from wind and sea waves and, because the atmospheric circulation responsible for wind and wind-driven waves is driven by solar heat, these are also forms of solar energy.

The outer layer of the Sun, which is what we see and the region from which the Sun radiates, is at a temperature of about 6000 K and it radiates energy at 73.5×10^6 W from every square metre of its visible surface (the photosphere; being entirely gaseous, the Sun has no solid surface). The figure can be calculated because the Sun behaves as a 'black body'. This is a body that absorbs all the energy falling on it and radiates energy at the maximum rate possible; the rate is calculated by using Stefan's law⁹ and is proportional to the absolute temperature raised to the fourth power.

The Sun radiates in all directions and the Earth, being a very small target at a distance of 150 million km, intercepts 0.0005 per cent of the total. At the top of the Earth's atmosphere this amounts to about 1360 W m⁻², a value known as the 'solar constant'.

Solar output is not as constant as this name suggests. Between 1981 and 1984, it decreased by 0.07 per cent (HIDORE AND OLIVER, 1993, p. 166). This is a small deviation, but a decrease of about 0.1 per cent sustained over a decade would be sufficient to produce major climatic effects and a 5 per cent decrease might trigger a major glaciation. Cyclical variations in the Earth's rotation and orbit also alter the solar constant. These are believed to be the major cause of large-scale climatic change, and variations in solar output, marked by changes in sunspot activity, are linked to less dramatic changes, such as the Little Ice Age, a period when average temperatures were lower than at present which lasted from about 1450 to 1880. Some scientists believe that the recent climatic warming and rise in atmospheric carbon dioxide concentration are both wholly due to the marked increase in energy output of the Sun since about 1966 (CALDER, 1999).

Radiant heat and light are both forms of electromagnetic radiation, varying only in their wavelengths, and the Sun radiates across the whole electromagnetic spectrum. According to Wien's law¹⁰, the wavelength at which a body radiates most intensely is inversely proportional to its temperature, so the hotter the body the shorter the wavelength at which it radiates most intensely. This is not surprising, because electromagnetic radiation travels only at the speed of light (beyond the Earth's atmosphere, in space, about 300000 km s^{-1}) and the only way its energy can increase is by reducing the wavelength. Very short-wave (high-energy) gamma (10^{-4} – $10^{-8} \mu\text{m}$) and X (10^{-3} – $10^{-5} \mu\text{m}$) solar radiation is absorbed in the upper atmosphere and none reaches the surface. Radiation with a wavelength between 0.2 and 0.4 μm is called 'ultraviolet' (UV); at wavelengths below 0.29 μm , most UV is absorbed by stratospheric oxygen (O_2) and ozone (O_3). The wavelengths between 0.4 and 0.7 μm are what we see as visible light, with violet at the short-wave end of the spectrum and red at the long-wave end. These are the wavelengths at which the Sun radiates most intensely, with an intensity peak at around 0.5 μm in the green part of the spectrum. It is the part of the spectrum to which our eyes are sensitive, for the obvious reason that the most intense radiation is also the most useful, although some animals have eyes receptive to slightly shorter or longer wavelengths. Beyond the red end of the visible spectrum lie the infra-red wavelengths (0.7 μm to 1 mm) and, with increasing wavelengths, microwaves and radio waves, the longest of which have wavelengths up to about 100 km.

The atmosphere is transparent to wavelengths longer than 0.29 μm , although water vapour absorbs energy in several narrow bands between 0.9 and 2.1 μm (BARRY AND CHORLEY, 1982, pp. 10 and 15). When radiant energy, as light or heat, strikes the surface of land or water its energy is absorbed and the surface is warmed. The Earth is not warmed evenly and Figure 2.10 shows how the energy is distributed. The equator faces the Sun, which is always directly overhead at noon. Consequently, its radiation is most intense at the equator. With increasing distance from the equator, the Sun is lower in the sky at all seasons and its radiation covers a larger area less intensely.

Although latitude is obviously important, and places in high latitudes tend to receive less solar energy than those in low latitudes, cloudiness modifies the general distribution quite strongly. The equatorial region does not receive the most intense insolation, because for much of the time clouds shade the surface, reflecting incoming sunlight. Tropical and subtropical deserts, where skies are mainly clear, receive 50 to 100 per cent more insolation than the equator and the dry interiors of North America and Eurasia are much sunnier than maritime regions.

Rather less than half of the solar radiation reaching the top of the atmosphere penetrates all the way to the surface. As Figure 2.11 shows, most of the 'lost' incoming radiation is reflected directly back into space, and about 10 per cent is absorbed or scattered by ozone, water vapour, and particulate matter in the troposphere.

It is scattering that gives the sky its colour. Radiation bounces off particles (mainly molecules) of a particular size in relation to its wavelength. All that changes is the direction of the radiation. There is no loss of energy, but shorter wavelengths scatter more than longer ones. This is called Rayleigh scattering, after Lord Rayleigh (1842–1919) who discovered it, and it reflects radiation in all directions. When the Sun is high in a clear sky, violet light is scattered and absorbed very high in the atmosphere and blue below it. Scattering diffuses the blue light evenly and so the sky appears blue. If the sky is hazy, dust particles scatter light of all wavelengths and the sky appears white. When the Sun is low, dust particles scatter light in the orange and red wavelengths, but shorter wavelengths are absorbed during the much longer passage of the light through the air, and the sky appears orange or red. Spherical particles larger than those responsible for Rayleigh scattering (more than about 0.1 μm) scatter light of all wavelengths, mainly without changing its direction. This is Mie scattering, discovered by Gustav Mie in 1908, and it tends to darken the

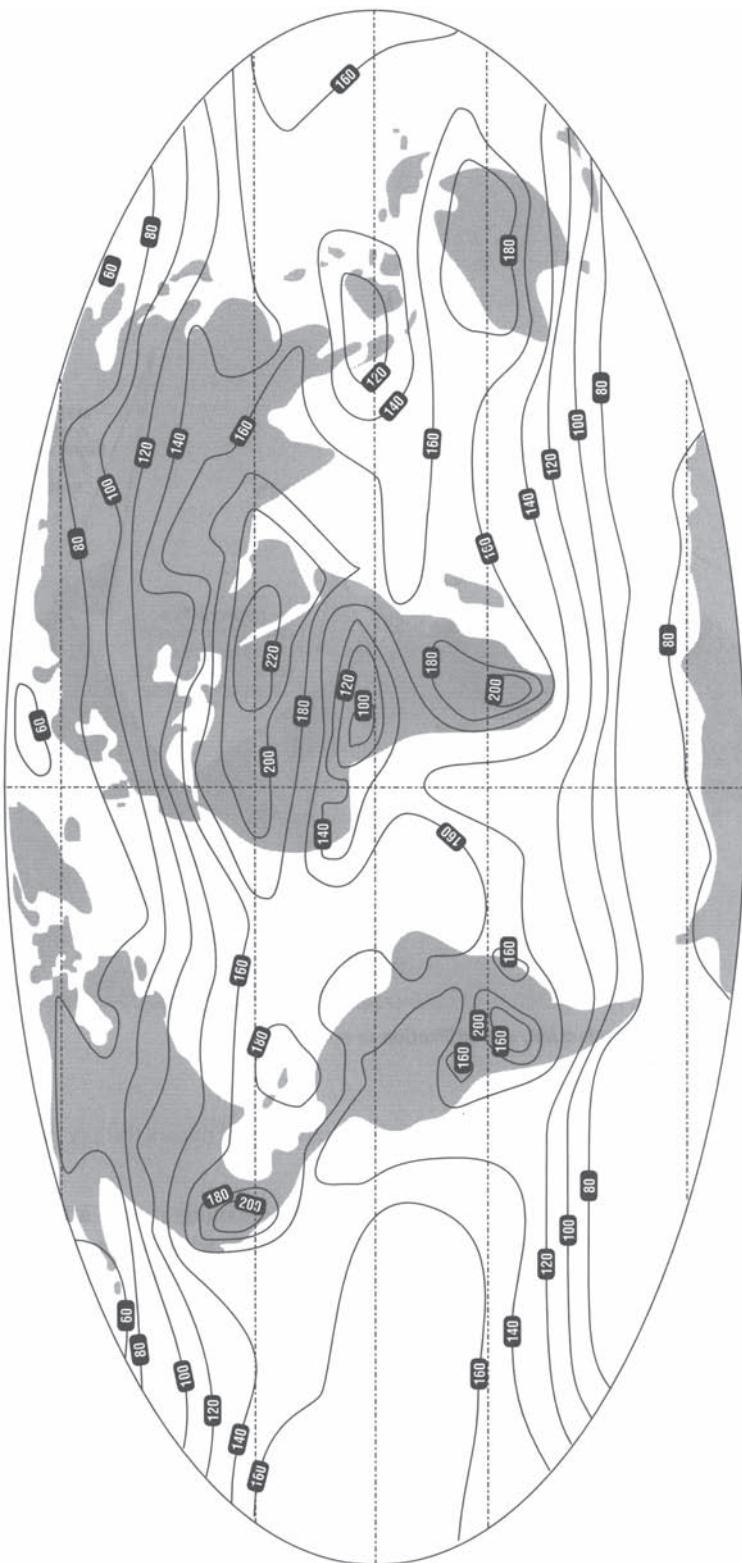


Figure 2.10 Average amount of solar radiation reaching the ground surface, in $\text{kcal cm}^{-2} \text{yr}^{-1}$ (1 $\text{kcal}=4186.8 \text{ J}$)

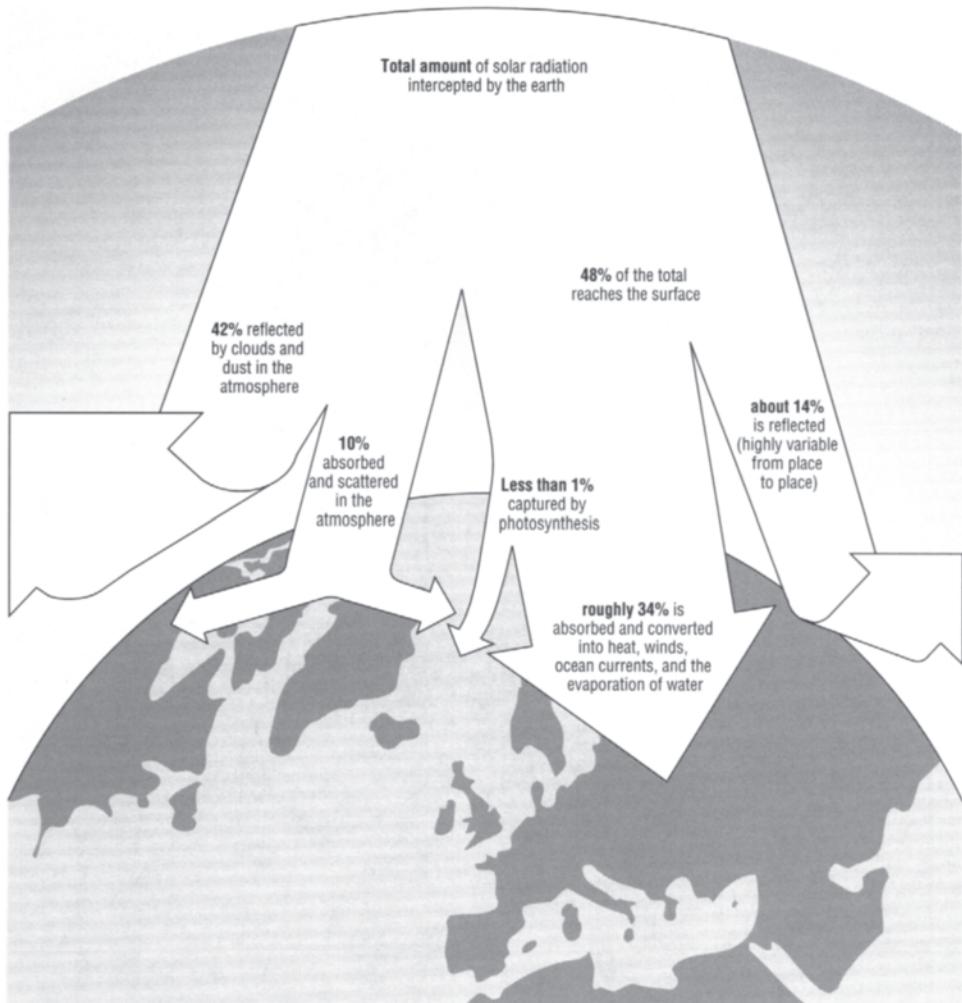


Figure 2.11 Absorption, reflection, and utilization of solar energy

sky colour by counteracting the effect of Rayleigh scattering; it makes the sky a darker blue after rain has washed out solid particles.

Once warmed, the Earth also behaves as a black body, radiating energy in the long, infra-red waveband. All the received energy is reradiated. All the portion which is captured by green plants and subsequently passed to animals that eat the plants is converted back into heat by the process of respiration and escapes from the Earth. This must be so, because if captured energy were retained permanently the Earth would grow continually hotter, and it does not. Overall, the amount of radiation received from the Sun is equal to the amount radiated into space from the surface of the Earth, but a proportion of the outgoing energy is retained for a time in the atmosphere. This produces the ‘greenhouse effect’.

Solar energy can be exploited for domestic and industrial use, as a so-called ‘renewable’ energy source, but none of the exploitative technologies is free from problems (RAVEN *ET AL.*, 1993, pp. 234–250).

Fast-growing crops, harvested to be burned, are being cultivated in several parts of the world as 'biomass' fuel. Willow (*Salix* species) and similar woody plants can be burned directly, after drying then chopping and compressing them, which reduces their bulk. Alcohols can be obtained from plants rich in sugar or starch and either used directly or dehydrated and mixed with gasoline to make 'gasohol'. Low petroleum prices led to a decline in the number of Brazilian cars being built to run on 'gasohol', but in 1999 car manufacturers announced an increase in production in an attempt to boost car sales. Fiat planned to raise output of these vehicles from 90 in August to 1,300 in September and Volkswagen planned an increase from 800 to 1,200. General Motors introduced a new model in September and Ford planned to relaunch its models in the spring of 2000. Methanol, an alternative liquid fuel, can also be obtained from plant material. Such fuels are renewable because they can be replaced easily by growing more of the fuel crop, and although they are based on carbon they make no contribution to the greenhouse effect, because the carbon they release when burned is absorbed photosynthetically by the plants which replace them; the carbon is recycled. Biomass crops occupy land, however, and if they are to be grown on the scale needed to supply useful amounts of fuel they could compete for space with food or fibre crops, and because they sell for a lower price than conventional crops they may be grown very intensively to maximize yields.

Solar heat is absorbed by a black surface. Manufactured solar collectors exploit this, using the absorbed heat to warm water, then transferring the heat to a hot water system. Collectors are limited geographically, because they are most efficient where insolation is greatest and do not work well in high latitudes. They have been installed on many buildings, usually attached to roofs or walls facing the Sun, but their high capital cost often makes the energy they provide more expensive than that supplied by the public utilities. In the tropics, however, direct solar heating can be used to distil water and for cooking, with real advantages.

Photovoltaic cells can be used in any latitude, because they convert light, not heat, into electrical power. In years to come this technology may provide useful amounts of energy, but at present its low efficiency (about 15 per cent) means very large arrays are required and the resulting power is much more expensive than electricity generated by other means. For many years, scientists and engineers have been discussing the technical feasibility of constructing truly vast arrays of photovoltaic cells in geostationary orbit and transmitting the power generated as microwaves to a receiving station on the surface, where they would be reconverted to electrical power. The amount of energy obtainable in this way would be great, although it would not be cheap, and at present no one can predict its environmental consequences.

Sunlight can be concentrated. A device developed in Israel by the commercial arm of the Weizmann Institute of Science and Boeing uses highly reflective mirrors (heliostats) to track the Sun and reflect sunlight to another reflector on top of a central tower. This reflector redirects the sunlight to a matrix of concentrators, which increase the intensity of the light 5000 to 10000 times, compared with the sunlight reaching the surface outside the facility. The concentrated light is fed to a receiver, called 'Porcupine' because it contains hundreds of ceramic pins arranged in a geometric pattern. Compressed air flowing across the pins is heated and channelled to gas turbines that generate electrical power. The prototype plant was installed late in 1999.

Wind power is also exploited widely, but it, too, suffers from the fact that although solar energy, as wind, is abundant, it is variable and very diffuse. The amount of energy captured by a wind turbine is proportional to the square of the diameter of the circle described by its blades and the cube of the wind speed (ALLABY, 1992, pp. 194–202) in a 32 km h⁻¹ wind, a 15-metre-diameter rotor linked to a generator operating at 50 per cent efficiency generates 24 kW of power. Most modern wind generators have a rated capacity of about 750 kW and are established in arrays ('wind farms'), each turbine occupying about 2 ha, the spacing necessary to avoid mutual interference, therefore up to 3000

turbines, occupying 6000 ha, are needed to match the output of a large conventional power station. The unreliability of the wind means conventional generating capacity must be held available for use when the wind speed is too low or so high that the blades must be feathered (turned edge-on to the wind) to stop the rotors turning. Suitable sites for such installations are limited, and in highly valued open landscapes they tend to be visually intrusive and arouse strong opposition. Were wind power to provide a substantial proportion of our energy requirements, there could be a risk that the very large installations might affect local climates by extracting a significant part of the energy of weather systems.

The vertical movement of sea waves can also be used to generate electricity. The technology is well advanced, but wave power suffers from disadvantages similar to those of wind power. The installations need to be large and both they and the cables carrying the energy they generate to shore must be able to withstand ocean storms. They must also be located in places where wave movement is large and reliable, but well clear of shipping lanes. This limits the availability of suitable sites. An alternative device, which occupies a much smaller area, extracts energy from the oscillation of waves within a cylindrical structure, and energy can also be obtained in still waters by exploiting the temperature difference between warm surface water and cold deep water.

Wind and wave power can probably be used most effectively on a small scale in places that are beyond the reach of conventional energy distribution systems, such as remote, sparsely populated islands where demand is modest and the cost of links to the mainland high. Direct solar energy, captured by solar collectors or photovoltaic arrays, may become more popular if they can be made more efficient and ways can be found to spread the high capital cost over the lifetime of the installation. Biomass conversion, exploiting energy captured by photosynthesis in green plants, is perhaps the most promising of the renewable technologies. It requires land that is surplus to rival agricultural needs; a criterion that perhaps is now being met in many parts of the European Union.

Compared with the energy our planet apparently receives from the Sun, the amount we derive from fossil and nuclear fuels seems trivial. It is tempting, therefore, to suppose that solar energy can be harnessed to provide environmentally benign power from an original source that is free. Unfortunately, the technical problems are formidable, the costs high, and the environmental consequences uncertain.

12 Albedo and heat capacity

Walk in a snow-covered landscape on a sunny day and you may feel more comfortable if you wear dark glasses. Indeed, you may be well advised to wear dark glasses, because the light may be bright enough to hurt your eyes. Once the snow has melted, and the ground is carpeted with plants, you will have less need of dark glasses. The light will not be so bright.

We can see objects because of the light reflected from them. Their colour is determined by the wavelengths of the light they reflect and their brightness by the amount. Freshly fallen snow reflects 80–90 per cent of the light falling on it, grass 18–25 per cent, and this is why you need dark glasses when crossing snow: it may be almost as bright as the Sun itself. Grass, on the other hand, is much duller.

The proportion of light reflected by a surface is called the ‘reflection coefficient’, or more usually ‘albedo’ of that surface. It can be measured and is usually expressed as a fraction or a percentage. As Table 2.1 shows, albedo varies widely from one surface to another.

Table 2.1 Albedos of various surfaces

Surface	Albedo (%)
<i>Snow</i>	
Fresh snow	80–90
Melting snow	40–60
Sea ice and old snow	30–40
<i>Cloud</i>	
Cumulonimbus cloud	70–90
Stratocumulus cloud	60–84
Cirrus cloud	40–50
<i>Dry land</i>	
Desert	25–30
Sand	30–35
Tundra	15–20
Grass	18–25
Dry savannah	25–30
Wet savannah	15–20
Deciduous forest	15–18
Coniferous forest	9–15
Tropical rain forest	7–15
Field crops	3–15
Dry ploughed field	5–25
Concrete	17–27
Asphalt	5–17
<i>Earth:</i>	
Overall	31
Land and water surface	14–16
Cloud	23

Table 2.2 Effect of the incident angle of radiation on water's albedo

Incident angle (°)	Albedo (%)
0	99 (or more)
10	35
30	6
50	2.5
90	2

About 70 per cent of the Earth's surface is covered by water. Its albedo varies according to the angle at which sunlight reflects from it. When the Sun is low in the sky much more light is reflected and the water looks brighter than when the Sun is high; in the latter case most of its light penetrates the water and is absorbed, and the water looks dark. Early and late in the day, when the water is calm, the occupants of open boats can develop sunburn quite quickly, even in cool weather. Table 2.2 compares the incident angle of radiation with the resultant albedo for water.

Reflected radiation does not warm the surface and so albedo has an important climatic effect, and one that can be modified by human intervention, although the relationship is more complex than it may seem. The clearance of tropical rain forest to grow field crops, for example, involves little change, but if that change is from forest to pasture for feeding livestock, the albedo could double. In this case, the ground would absorb less heat, so there would be less evaporation of water and less cloud would form. This would reduce the average cloud albedo, however, so increasing the amount of radiation reaching the surface and warming it again, evaporating more water and increasing cloudiness once more, but not necessarily to its original value.

This rather intricate relationship illustrates an important point. Climate is strongly subject to feedback effects. In most cases, as in our example, these tend to stabilize conditions, as negative feedback, but positive feedback also occurs. It exaggerates effects and so has a destabilizing effect which can be felt rapidly, as in the onset of glaciation and glacial melting. Eventually, positive feedback is overridden by negative feedback and a destabilized system finds a new level of stability.

Clean air reflects no light, but air is seldom clean. It contains very small particles, called 'aerosols'. In the upper troposphere and at lower levels over the open sea, concentrations range from about 100 to 600 per cm³, but at low level over continents they are much higher and in industrial regions can reach millions. They vary in size

from 10⁻³ to 10² µm and gravity has little effect on them, because they are so small. They tend to reflect short-wave radiation, thus increasing planetary albedo, and in saturated air water vapour will condense on aerosols smaller than about 0.5 µm, encouraging cloud formation and also increasing albedo (and removing the aerosols in precipitation). Increasing the albedo has a cooling effect, but aerosols absorb radiation at infra-red wavelengths, so those in the lower atmosphere also have a warming influence; for those reaching the stratosphere, on the other hand, the cooling influence of increased albedo is dominant. To complicate matters further, aerosols settling on to clean snow 'dirty' it, reducing its albedo.

Aerosols are released into the air by volcanoes, as salt crystals formed when droplets of sea spray evaporate, from forest fires, and as tiny soil particles raised by the wind. They are also produced by a range of human activities, especially the burning of fuels. It is difficult to separate natural sources from those linked directly to human activities, and both vary from season to season and year to year but, on average, agriculture and industry account for about one-third of the particulate matter in the air. Forest clearance and the overgrazing of marginal land in semi-arid regions, for example, leads to large injections of small particles as wind-blown soil.

From time to time people suggest altering albedos to trigger climatic change. Particles injected into the upper troposphere might increase the formation of cirriform cloud, for example, and particles injected into the stratosphere might increase planetary albedo for several years. In both cases the quantities of particles required would be huge. It might also be possible to reduce albedo in deserts, by colouring large areas black (perhaps by covering them with plastic sheeting). Such ‘thermal mountains’ would stimulate convection, hopefully leading to the formation of cumuliform clouds that would release rain. Most climatologists are wary of such schemes, suspecting that in the unlikely event that they worked the unanticipated consequences might be unpleasant. Happily, perhaps, their high cost makes them unattractive to governments.

Varying albedo means some surfaces absorb more solar energy than others, but there are also wide differences in the way heat is absorbed. On a really hot summer day the surface temperature of sand on a beach may be high enough to make it painful to walk across it in bare feet, whereas the water is cool, yet both sand and water are exposed to the same amount of insolation. Dig your feet into the sand, however, and you soon reach a cooler level. The differences were measured over a Saharan sand dune at 1600 hours, when the temperature reached its maximum. The air temperature was a little over 40°C, that of the sand surface 65°C, but 30 cm below the surface the temperature was about 38°C and 75 cm below the surface it was 25°C (BARRY AND CHORLEY, 1982, p. 288). Soon after sunset, of course, the sand surface would feel cool.

Different materials vary in their response to radiant energy because they have different heat capacities. Heat capacity is calculated as the ratio of the amount of energy applied, to the resulting rise in temperature. The heat capacity of water is much greater than that of rock. This means that much more energy is needed to raise the temperature of water than of rock, or any substance made from rock. It also means that water loses heat much more slowly than rock (HIDORE AND OLIVER, 1993, p. 58). Consequently, water responds to insolation by warming and cooling slowly and land by warming and cooling quickly. This explains the difference in temperature between the sand on a beach and the water beside it, but it also has profound climatic implications.

The rate at which temperature decreases below the surface depends on the conductivity of the material and its mobility, which affects the transfer of heat by convection. Sand grains conduct heat poorly, which is why a layer of cool sand lies at quite shallow depth. Although water is not a very good conductor of heat, heat moves through it readily by convection, and turbulence due to the wind mixes warmed surface water with cooler water immediately beneath it.

13 The greenhouse effect

Since the solar constant is known it is possible to calculate what the ‘black-body’ temperature of the Earth should be: it is 250 K (-23°C) (HARVEY, 1976, pp. 43–44). This is what the average temperature at the surface would be were it not for the absorption of long-wave radiation by the atmosphere, which delays the loss of heat and thus warms the planet. The absorption of radiation modifies the

climate, or ‘forces’ it into a (warmer) state than it would be otherwise. The actual average surface temperature is about 288 K (+15°C), a difference of 38°C. The forcing that achieves this difference is called the ‘greenhouse effect’ and, clearly, without it life on Earth would be very uncomfortable, if it were possible at all with surface water, including the oceans to considerable depth, frozen solid and the liquid water beneath the ice markedly more saline because it would contain the salt removed from solution as water crystallized to form ice.

Nitrogen and oxygen are almost completely transparent to electromagnetic radiation at wavelengths greater than 0.29 μm, but some of the incoming solar radiation is absorbed by other constituents of the atmosphere. Of the total, about 4 per cent is absorbed by stratospheric ozone, 20 per cent in the infra-red band by carbon dioxide and 13 per cent by water vapour (in three narrow infra-red wavebands at about 1.5 μm, 2.0 μm, and 2.5–4.5 μm), and 6 per cent by water droplets and dust. At wavelengths greater than about 4.0 μm, however, several atmospheric gases absorb radiation, each in particular wavebands related to the size of its molecules. The significance of this arises from the fact that the Earth, warmed by the Sun, behaves like a black body at a temperature of 288 K and emits electromagnetic radiation at 4–100 μm, with a peak of intensity around 10 μm. More than 90 per cent of this outgoing long-wave radiation is absorbed in the atmosphere (BARRY AND CHORLEY, 1982, pp. 33–35). The remainder, about 6 per cent, with wavelengths between those at which it can be absorbed, escapes into space. These ‘gaps’ in the absorption bands, at about 8.5 μm and 13.0 μm, are called the ‘atmospheric window’.

Molecules which absorb radiation reradiate it in all directions (see Figure 2.12). Some returns to the surface, some is absorbed by other atmospheric molecules and some is radiated upwards, out into space. Of course, the surface is warmed by the Sun only during daytime, but its heat is radiated away by night as well as by day. Eventually, as much energy leaves the Earth as reaches it from the Sun. It must do, because otherwise the atmospheric and surface temperatures would either rise or fall steadily over time; the overall energy budget must balance, and it does, although the transfer of energy is complicated.

The ‘greenhouse’ metaphor is colourful but a little misleading. It is true that the glass of a greenhouse is transparent to short-wave radiation and partly opaque to infra-red radiation, so its action is similar to that of the absorbing gases in the atmosphere, but the temperature difference inside and outside a greenhouse is due mainly to the fact that air inside is prevented from being cooled by mixing with air outside. With this minor qualification, however, the

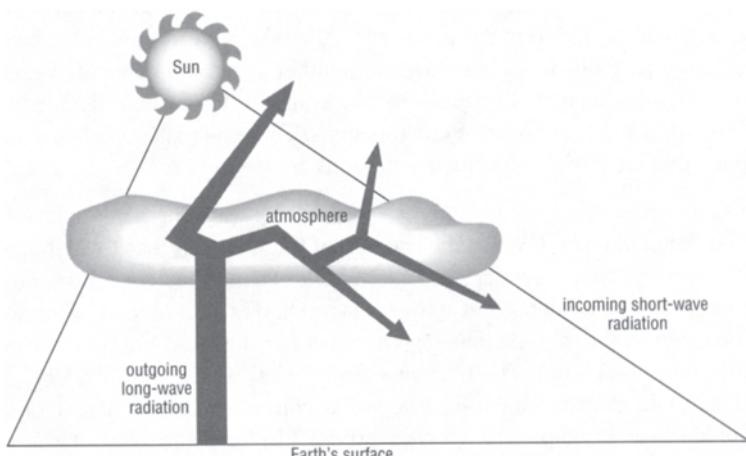


Figure 2.12 *The greenhouse effect*

atmospheric greenhouse effect is real and important, and the gases which cause it are justly known as ‘greenhouse gases’.

Both the global climate and atmospheric concentrations of greenhouse gases vary from time to time. Studies of air trapped in bubbles inside ice cores from Greenland and from the Russian Vostok station in Antarctica have revealed a clear and direct relationship between these variations and air temperature, in the case of the Vostok cores back to about 160 000 years ago. The correlation is convincing, although it is possible that the fluctuating greenhouse-gas concentration is an effect of temperature change rather than the cause of it. As temperatures rose at the end of the last ice age, the increase in the atmospheric concentration of carbon dioxide lagged behind the temperature (CALDER, 1999) and so carbon dioxide cannot have been the cause of the warming. There is also evidence that the carbon dioxide concentration was far from constant prior to the start of the Industrial Revolution (WAGNER *ET AL.*, 1999). Carbon dioxide measurements taken from air bubbles trapped in ice cores are unreliable, because carbon dioxide is soluble in solid ice.

Nor has the temperature always been linked to the concentration of carbon dioxide. The two were uncoupled between about 17 and 43 million years ago. The air then contained less than two-thirds of the present concentration of carbon dioxide ($180\text{--}240 \mu\text{mol mol}^{-1}$ compared with $360 \mu\text{mol mol}^{-1}$ today), but the climate was up to 6°C warmer than it is today (COWLING, 1999).

Nevertheless, it is estimated that the atmospheric carbon dioxide concentration immediately prior to the Industrial Revolution was about $280 \mu\text{mol mol}^{-1}$ and that the increase since then has been due entirely to emissions from the burning of fossil fuels. This may not be the case. The solubility of gases, including carbon dioxide, is inversely proportional to the temperature. A rise in temperature, therefore, will cause dissolved carbon dioxide to bubble out of the oceans. This is called the ‘warm champagne’ effect. Rising temperature will also stimulate aerobic bacteria. Their respiration will release carbon dioxide. This is called the ‘warm beer’ effect (CALDER, 1999).

Carbon dioxide is the best-known greenhouse gas, because it is the most abundant of those over which we can exert some control, but it is not the only one. Methane, produced naturally, for example by termites, but also by farmed livestock and from wet-rice farming (present concentration about 1.7 ppm), nitrous oxide (0.31 ppm) and tropospheric ozone (0.06 ppm), products from the burning of fuels in furnaces and car engines, and the industrially manufactured compounds CFC-11 (0.00026 ppm) and CFC-12 (0.00044 ppm) are also important. The most important of all, however, is water vapour. This enters into the calculations only indirectly, because its concentration varies greatly from place to place and from day to day and because it is strongly affected by temperature. Its influence, therefore, tends to add to those of the other gases and generally varies as they do. Figure 2.13 shows the anticipated changes in concentration for carbon dioxide, methane, and CFC-12, which is one of the family of CFC compounds. These increases are based on the (uncertain) assumption that industrial and vehicle emissions are the only source of carbon dioxide.

All greenhouse-gas effects are usually expressed as ‘global warming potentials’ (GWPs) which relate them to carbon dioxide. GWPs take account of the wavelengths at which particular molecules absorb, some of which overlap, and the length of time they remain in the atmosphere before decomposing or being deposited at the surface. On this basis, over a 100-year period, with carbon dioxide given a value of 1, methane has a value of 11 (i.e. it is 11 times more effective than carbon dioxide, molecule for molecule), nitrous oxide 270, CFC-11 3400, and CFC-12 7100. The estimates of future climatic warming are based on the consequences calculated for a doubling of the carbon dioxide concentration, which includes the GWPs for all the relevant gases. Figure 2.14 shows that, depending on the sensitivity of the atmosphere to greenhouse forcing, a doubling of carbon dioxide would raise the average global temperature by $1.5\text{--}4.5^\circ\text{C}$, with a ‘best estimate’

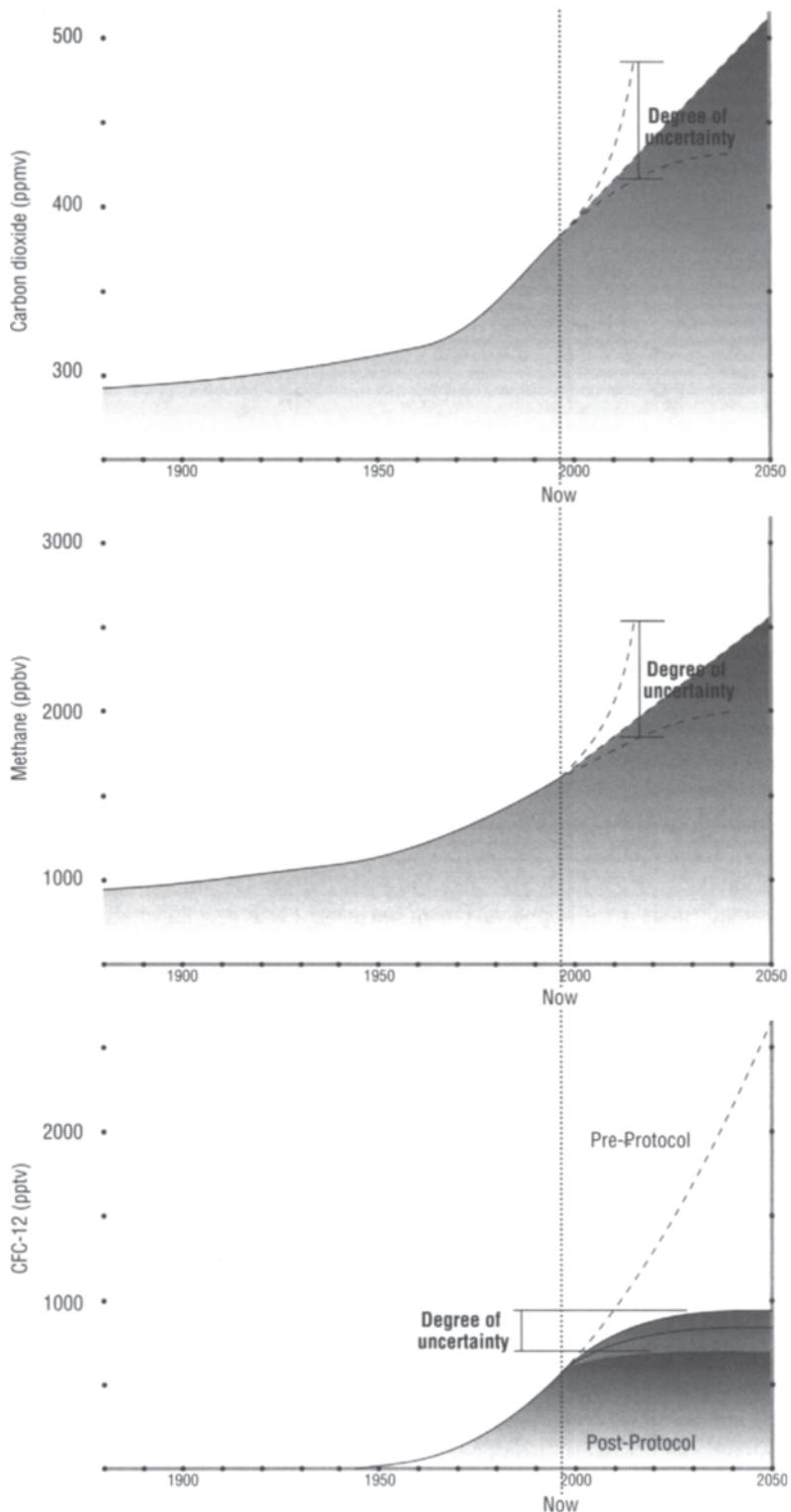


Figure 2.13 Anticipated changes in concentration of three greenhouse gases

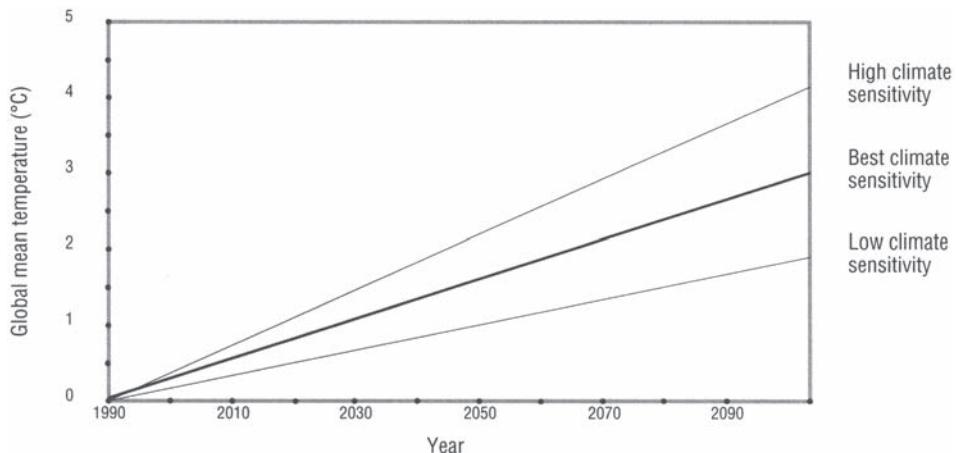


Figure 2.14 IPCC estimates of climate change if atmospheric CO₂ doubles

of 2.5°C, and at the current rate of increase in greenhouse-gas concentrations these temperatures would be reached by around 2100. During the same period, warming of the oceans is calculated to cause them to expand, producing a rise in sea level of 2–4 cm per decade (IPCC, 1992).

It is not certain that sea levels have risen, although predictions of global warming include an assertion that they have risen world-wide by about 25 cm over the past century. In 1841, the explorer Sir James Clark Ross visited Tasmania, where he met Thomas Lempriere, an amateur meteorologist. The two men installed an Ordnance Survey Bench Mark, chiselling it into a rock face at a place called the Isle of the Dead, near Port Arthur. It was positioned with great care and precision and was meant to act as a sea-level gauge. The gauge has been rediscovered and the Tasmanian climatologist John L.Daly visited it in August 1999. He found that it remains visible above the water line, despite the supposed rise in sea level. It is uncertain whether the gauge was set at a level close to the high tide mark or at the mean tide level. If it was close to the high tide level it shows that sea level has not changed since 1841. If it was at the mean tide level it shows the sea level has fallen. (www.vision.net.au/~daly/)

The mean global temperature increased by 0.37°C between 1881 and 1940. The temperature fell from 1940 until the 1970s, since when it has risen again, but there is no clear evidence of any warming between 1980 and 1998. The total warming from 1881 to 1993 amounts to 0.54°C. Two-thirds of the warming occurred before 1940—and before the main rise in the atmospheric carbon dioxide concentration—and 1881 was an unusually cold year (BALLING, 1995).

Mean temperatures are calculated from three sets of data. Weather stations and ships record surface temperatures, balloon sondes record temperatures above the surface, and TIROS-N satellites operated by NASA on behalf of the National Oceanic and Atmospheric Administration (NOAA) measure temperatures from space (www.atmos.uah.edu/essl/msu/background.html). Surface measurements are difficult to interpret over long periods. This is because weather stations that are established in open country may gradually be affected by nearby urban development and road building, which will raise the temperature by an ‘urban heat island’ effect, producing an illusory warming trend. There is also the possibility that, over the years, changes in staff may lead to unrecorded changes in the time of day when measurements are made. Ships measure the temperature of sea water below the surface, but different ships do so at different levels. Their measurements of air temperatures are unreliable for similar reasons. Over the course of this century, ships have become larger, so their decks from

which temperatures are measured are higher above the sea than they used to be. In any case, ships' thermometers are not calibrated against a standard.

Balloon readings are much more reliable. Weather balloons are released twice every day, usually at noon and midnight Greenwich time, from about 1000 sites. These sites are located predominantly in industrial countries, however, so records derived from them may not be typical of the world as a whole.

Satellite measurements are by far the most reliable of all. More than 30000 measurements are made every day.

Measurements from surface stations show temperatures are rising by 0.15°C per decade from 1979 to 1997. This is a much smaller increase than the IPCC 'best estimate' of 2.5°C . Over the same period, NOAA weather balloons show temperatures are falling by 0.07°C per decade and balloon data from the UK Meteorological Office show them falling by 0.02°C per decade. The satellite measurements show them falling by 0.01°C per decade (science.nasa.gov/newhome/headlines/notebook/essd13aug98_1.htm). The mean temperature measured by satellite since January 1998 has risen by about 0.04°C per decade. This warming was caused by the very strong El Niño event in 1998. The warming is still very much smaller than that predicted by the IPCC.

The effects of climate forcing are being studied by teams in several countries and their results are drawn together into a scientific consensus by the IPCC. This body, involving hundreds of specialists from all over the world, was established in 1988 by the World Meteorological Organization and the UN Environment Programme to advise governments. By no means all climatologists agree with the IPCC conclusions, however.

Governments became involved following a meeting held at Villach, Austria, in 1985 under the auspices of the International Council for Scientific Unions (ICSU, now called the International Council for Science). There, research scientists, including ecologists and experts on climate and energy-demand modelling, concluded that global warming was a real threat and more research was needed and, supported by environmentalist groups, the topic quickly acquired political influence. This politicization and resultant popular dramatization of a very complex and uncertain issue has attracted criticism (BOEHMER-CHRISTIANSEN, 1994; BOLIN, 1995).

Studies of climate forcing begin with estimates of ways in which the chemical composition of the atmosphere may change in the future, to produce an 'emissions scenario'. This requires a knowledge of the sources from which greenhouse gases are released, the sinks into which they are absorbed, and ways the sinks may respond to increased loading. The oceans are the most important sink for carbon dioxide, but the behaviour of the sinks is incompletely understood and no sink has been identified for a significant fraction of the carbon dioxide known to have been emitted. Measurements of greenhouse-gas concentrations must also distinguish between genuine changes, the 'signal', and natural variations, the 'noise'. Carbon dioxide levels vary seasonally, for example, in response to the growing season for plants.

General circulation models (GCMs) are then constructed. These are based on a notional three-dimensional grid placed over the entire Earth. Atmospheric behaviour is calculated according to physical laws for every grid intersection. The input data for each calculation include the state at adjacent grid points as well as data introduced by the modeller, and so they trace the evolution of the atmosphere, simulating the climate. Using the known present state of the atmosphere, the model is used to simulate the climate over several decades and its results compared with actual climate records. If this test proves satisfactory, changes in atmospheric composition, based on the emissions scenario, are introduced to the model and their consequences evaluated.

Modelling on this scale requires massive computing power. Even when the fastest supercomputers are used, the grid must be fairly coarse to keep the number of intersection points to a manageable level. This means that some phenomena, such as cloud formation, must be greatly simplified, because they occur on a smaller scale than the $100 \times 100 \times 10$ km grid boxes. Most GCMs make similar simplified allowances for the mixing of surface and deep ocean water, although the latest ‘coupled’ models (CGCMs) treat the oceans as of complexity comparable to that of the atmosphere.

GCMs are being improved constantly and scientific understanding of atmospheric and oceanic processes is increasing rapidly, but much remains to be learned and estimates of the regional consequences of a general warming vary widely. The IPCC finds, for example, that warming will be reduced by 60 per cent or more over the northern North Atlantic and around Antarctica. The best illustration of the uncertainties surrounding the calculations centres on water vapour. If the temperature rises, more water will evaporate. Water vapour is a greenhouse gas, so this will accelerate the warming trend, but a more humid atmosphere will also be a cloudier atmosphere. As water vapour condenses to form cloud the latent heat of condensation, released into the surrounding air, also has a warming effect, but clouds themselves may have either a warming or a cooling effect. Generally, clouds at low level have a high albedo and, therefore, cool the surface, while high-level clouds absorb radiation and have a warming effect. Most GCMs predict an increase in middle- and high-level cloud, with a consequent warming effect, but an increase in cloud amount, with deeper cloud cover, might reduce warming. It is important to know how much cloud will form and its type, but at present this cannot be calculated for given atmospheric conditions.

Many climatologists accept there is a real possibility of global climatic warming due to an enhanced greenhouse effect. As with many large-scale changes, there would be winners and losers. Were climate belts to shift toward higher latitudes, which seems the most likely overall result, parts of the Sahara and southern Russia would receive increased rainfall. They would benefit and their agricultural output would increase. On the other hand southern Europe and the United States cereal belt might become drier. If warming produced a rate of evaporation that exceeded the increase in the rate of precipitation, soils would become more arid. It may be, however, that warming will be experienced as a reduction in the fall of temperature at night, due to increased cloudiness, with little or no change in daytime temperatures. In that event, nighttime frosts would become less frequent, soils would become somewhat moister, and agriculture would benefit.

Environmentalists favour the ‘precautionary principle’. This holds that if there is a chance of adverse change we should not wait until the risk can be scientifically confirmed before taking action to minimize it. When world leaders agreed to reduce greenhouse-gas emissions at the 1992 UN Conference on Environment and Development (the ‘Rio summit’) they did so in accordance with this principle. There are critics of the principle, however, who point to the cost and difficulty of pursuing policies that may prove unnecessary. The principle holds that if any innovation appears to entail a risk of serious or irreversible harm to human health or the environment, then precautionary measures should be taken to avert that risk. This seems to obtain even if the link between the innovation and the harm has not been proven or if it is weak and the harm is unlikely to occur. It can be argued that this weighs the possible advantages and disadvantages of innovation, but loads the scales in favour of the disadvantages. This could have a paralyzing effect if the only way to determine whether the risk is genuine is to undertake the innovation—which the possibility of risk forbids (HOLM AND HARRIS, 1999). Not surprisingly, others disagree, maintaining that the principle does not necessarily prohibit innovation but does encourage preventive action in the face of uncertainty (RAFFENSPERGER *ET AL.*, 1999). In this area, as in many others, the environmental science is uncertain and its translation into political action far from simple.

14 The evolution, composition, and structure of the atmosphere

When the Earth first formed, it may have had a thin atmosphere of light gases, mainly hydrogen and helium, derived from the stellar nebula. If so, this atmosphere was lost when the Sun began to radiate, supplying the gas molecules with the energy they needed to escape the planet's gravitational attraction. Volcanism then released the gases which formed a new atmosphere. No one knows the precise composition of that atmosphere, but probably it was rich in carbon dioxide and contained only traces of nitrogen and free oxygen. Our present atmosphere has evolved to its present state, partly (some would say mainly) as a result of biological processes.

These processes continue to maintain it. Nitrogen, for example, is chemically somewhat inert but, in the presence of oxygen and with a sufficient application of energy, it will oxidize to nitrate (NO_3). This reacts with water to form nitric acid (HNO_3) and is washed to the ground. Lightning supplies enough energy for the oxidation, and in the world as a whole it is estimated that there are about 1800 thunderstorms, with lightning, happening at any one time, delivering some 100 million tonnes of fixed nitrogen to the surface every year. At this rate it would not take long to strip the atmosphere of its nitrogen, were it not for the activities of denitrifying bacteria, which utilize nitrogen compounds in the soil and release gaseous nitrogen as a metabolic by-product. Were the air to be seriously depleted of nitrogen the proportion (partial pressure) of oxygen would increase and were it to increase to more than about 25 per cent exposed carbon compounds would burn even more readily than they do. This would reduce the amount of oxygen, replacing it with carbon dioxide. As it is, carbon dioxide is removed from the air by green plants as well as by dissolving in rain water.

To a considerable extent, therefore, our present atmosphere has been constructed and is maintained by living organisms. It is an essential component of the environment. It is also the route by which many organisms are disseminated and most nutrients cycled.

Air is a mixture of gases. They are not combined, so air itself is not a chemical compound, although some of its minor constituent gases are.

As the many practical uses for compressed air (in car and bicycle tyres, for example) testify, air is highly compressible. The total weight of the atmosphere (with a mass of about 5×10^{15} tonnes) compresses its lower layers, so the average density of the air decreases from about 1.2 kg m^{-3} at sea level to about 0.7 kg m^{-3} at a height of 5 km. Half of the total mass of the atmosphere lies below 5 km. The pressure exerted by the atmosphere is usually expressed in millibars (mb) in weather forecasts or in pascals (Pa) by physicists ($1 \text{ mb} = 100 \text{ N m}^{-2} = 100 \text{ Pa}$). The average sealevel pressure is 1013.2 mb (or 101320 Pa).

One consequence of its compressibility is that for all practical purposes the atmosphere is quite shallow. Although it has no clearly defined upper boundary, merging imperceptibly with the solar atmosphere at an altitude of about 80000 km (ALLABY, 1992, p. 50), pressure and therefore density decrease logarithmically with height. At 30 km, air density is at only 0.02 per cent of its sea-level value. The part of the atmosphere in which weather occurs is confined to the lowermost 8–16 km. None of the familiar meteorological phenomena occur above this height, but they are influenced by events in the lower reaches of the overlying layer, up to a height of about 30 km.

The air is warmed convectively from below, by contact with the surface. As it warms it expands and as it expands it cools. Temperature and density determine the amount of water vapour air can hold (its humidity), and the combination of the humidity, density, and temperature imposes a layered structure on the atmosphere. Figure 2.15 illustrates this structure, relating it to height, temperature, and pressure.

The equation of state

Boyle's law relates the volume of a gas to its pressure ($P=k_1/V$, where P is pressure, V is volume, and k_1 is a constant) and Charles's law relates pressure and temperature ($V=k_2T$, where T is absolute temperature and k_2 is another constant). Combined, these produce the 'equation of state': $P=\rho RT$, where P is pressure, ρ is density, T is absolute temperature, and R is the gas constant: $8.314 \text{ J K}^{-1} \text{ mol}^{-1}$.

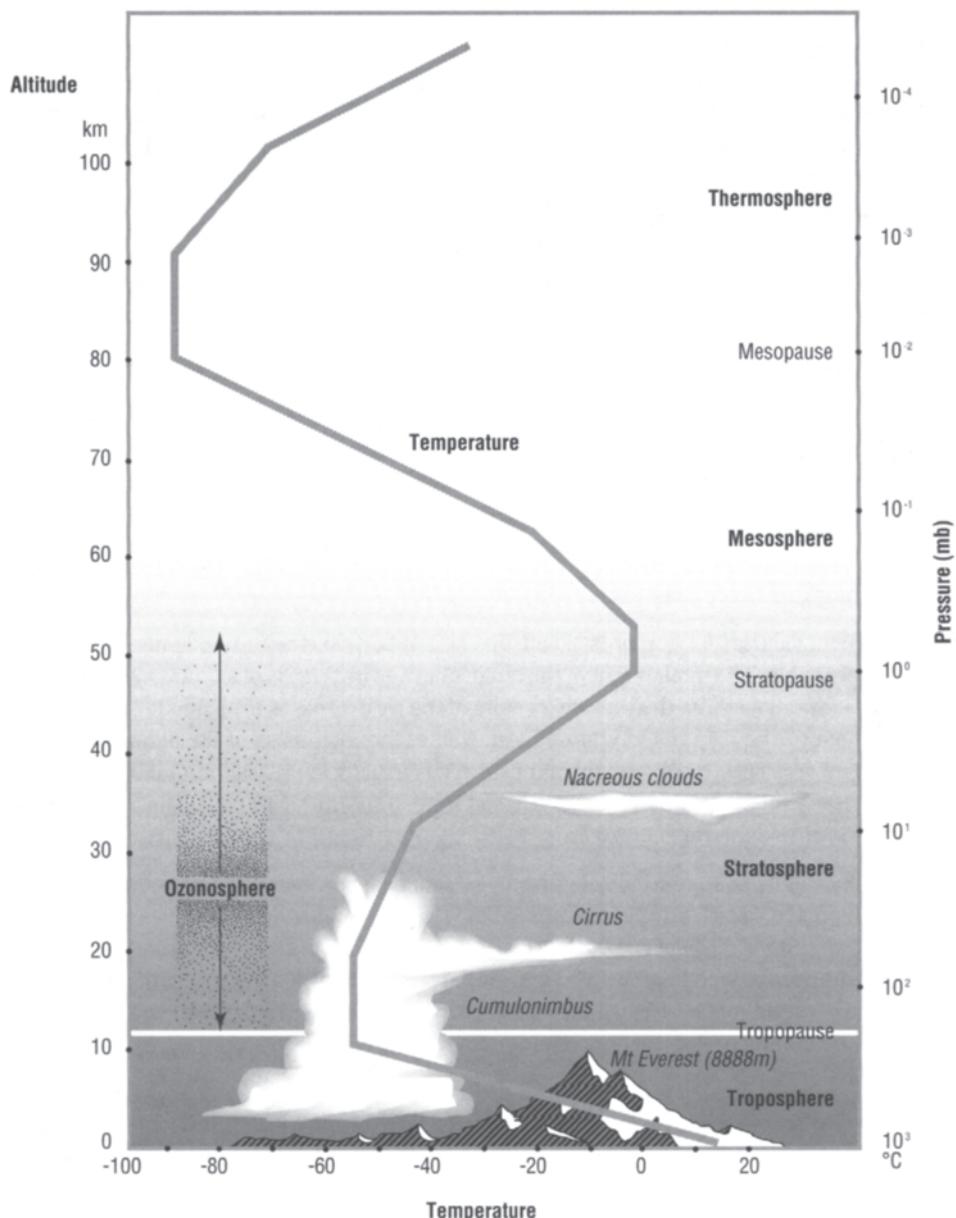


Figure 2.15 Structure of the atmosphere

The lowest layer, the troposphere, extends from the surface to an upper boundary, the tropopause, the height of which varies, but averages about 16 km at the equator and 8 km at the poles. Within the troposphere temperature decreases with height by an average of about $6.5^{\circ}\text{C km}^{-1}$ (called the ‘lapse rate’). Above the tropopause temperature remains constant with height to about 20 km. Rising tropospheric air is trapped below the region of constant temperature, which forms a permanent temperature inversion, a layer in which temperature remains constant or increases with altitude, rather than falling. It is this inversion which confines meteorological phenomena to the troposphere.

Temperature then increases from a minimum of about -80°C at the equator in summer, when the tropopause is at its highest, to 0°C or even higher at about 50 km. This region is the stratosphere and its upper boundary is the stratopause. In the mesosphere, above the stratopause, temperature once more decreases with height, to about -90°C at the mesopause, about 80 km, then rises again through the thermosphere. At about 350 km the temperature may exceed 900°C , probably because of the energy imparted by absorption of ultraviolet radiation by atomic oxygen, but the air is so rarefied that objects such as satellites are not warmed by it, although it still exerts measurable drag on spacecraft moving through it.

Between about 30 and 60 km the density of oxygen molecules is high enough to intercept most of the incoming solar ultraviolet radiation at wavelengths below $0.29\text{ }\mu\text{m}$. The energy imparted to them separates the molecules ($\text{O}_2 \rightarrow \text{O} + \text{O}$). Some of the oxygen atoms then combine with oxygen molecules to form ozone ($\text{O} + \text{O}_2 \rightarrow \text{O}_3$). Ozone is unstable and may decompose either by encountering more oxygen atoms ($\text{O}_3 + \text{O} \rightarrow 2\text{O}_2$) or by absorbing more ultraviolet radiation. Ozone is, therefore, constantly forming, decomposing, and re-forming, and the process is in equilibrium above about 40 km. There is also some transport of ozone from low to high latitudes. There is some mixing of stratospheric air, however, as a result of which a small amount of ozone is transported downward, to accumulate between about 20 and 25 km. This is the ‘ozone layer’. Its density varies, being lowest over the equator and high over latitudes above 50° . Ordinarily, ozone levels are also high over polar regions in early spring. This is because ozone is neither formed nor destroyed during the polar night, when there is no radiation to drive the reactions, and ozone transported from lower latitudes is stored (BARRY AND CHORLEY, 1982, pp. 2–3). Despite being known as the ‘ozone layer’, if the air at that altitude were compressed to sea-level pressure the ozone would contribute only about 3 mm to it. In itself, the ozone layer does not shield the surface from ultraviolet radiation, but indicates that radiation is being absorbed at a greater height, shielding both the surface and the ozone layer.

The thickness of the ozone layer is often reported in Dobson units (DU). This unit was devised by G.M.B.Dobson, a British physicist who studied stratospheric ozone in the 1920s. It refers to the thickness of the layer that a gas would form if all the other atmospheric gases were removed and the gas in question were subjected to standard sea-level pressure. In the case of ozone, 1 Dobson unit corresponds to a thickness of 0.01 mm and the amount of ozone in the ozone layer is typically 220–460 DU, corresponding to a layer 2.2–4.6 mm thick.

The depletion of the ozone layer over Antarctica, first observed in 1986 (ALLABY, 1992, pp. 159–161) by a British scientist, occurs just as spring is commencing. During the polar night, a vortex of very still air forms over Antarctica within which the temperature may be as low as -84°C . Clouds of ice crystals, called polar stratospheric clouds (PSC), form inside the vortex and a series of chemical reactions on the surface of the ice crystals results in chlorine monoxide (ClO) combining to form Cl_2O_2 . These molecules break down when exposed to sunlight ($\text{Cl}_2\text{O}_2 \rightarrow 2\text{Cl} + \text{O}_2$) and the chlorine atoms combine with ozone in two steps which release free chlorine once more to repeat the process so that a single chlorine atom can destroy many

thousands of ozone molecules ($\text{Cl} + \text{O}_3 \rightarrow \text{ClO} + \text{O}_2$; $\text{ClO} + \text{O} \rightarrow \text{Cl} + \text{O}_2$) (HIDORE AND OLIVER, 1993, pp. 74–77). CFCs are believed to be the principal source of stratospheric chlorine. Although very stable, they are decomposed by ultraviolet radiation at wavelengths below 0.23 μm, releasing free chlorine. As spring advances, the vortex disappears, ozone moves poleward from lower latitudes, and the ozone layer recovers. Seasonal depletion over the Arctic has also been reported, but it is less severe and of shorter duration, because Arctic winter stratospheric temperatures are higher than those of the Antarctic and a polar vortex rarely forms.

Ozone depletion may lead to increased exposure to ultraviolet radiation at the surface, the biological significance of which is uncertain. Ultraviolet radiation causes cataracts and non-melanoma skin cancer in fair-skinned humans (recent increases being due to the popularity of sunbathing in hot climates to which people are not acclimatized, and not to ozone depletion). It might have an adverse effect on land plants especially susceptible to it and may also affect organisms living in the uppermost few millimetres of the ocean surface; below that depth ultraviolet radiation is absorbed by sea water.

Ozone is a very minor constituent of the atmosphere. In the troposphere it occurs locally, some naturally but more commonly as a pollutant which causes respiratory irritation in humans and can damage plants, produced by photochemical reactions involving vehicle exhaust fumes. It is a constituent of photochemical smog and responsible for some of the damage attributed to acid rain.

The principal ingredients of the atmosphere are listed in Table 2.3. Water vapour comprises up to 4 per cent in the lower atmosphere, but above about 12 km it is virtually absent. The source of the water vapour from which PSCs form is unknown; it may have entered the stratosphere as water vapour and accumulated in the polar vortex or may result from the oxidation of methane ($\text{CH}_4 + 2\text{O}_2 \rightarrow \text{CO}_2 + 2\text{H}_2\text{O}$). Water vapour apart, the composition of the atmosphere remains constant to a considerable height, because of mixing caused by turbulence. Beyond the mesosphere, however, the proportions of its ingredients change. Figure 2.16 illustrates how the chemical composition changes with height.

Table 2.3 Average composition of the troposphere and lower stratosphere

Constituent	% by volume
Nitrogen	78.08
Oxygen	20.94
Argon	0.93
Carbon dioxide	0.035
Neon	0.0018
Helium	0.0005
Ozone	0.00006
Hydrogen	0.00005
Methane	0.00017
Krypton	trace
Xenon	trace

15 General circulation of the atmosphere

If the Earth faced it directly, the Sun would be overhead at the equator at noon every day of the year. This would have a profound effect on climates, because there would be no seasons. In fact, of course, the Earth is tilted on its axis, so we do not face the Sun directly. Our orbit traces the circumference of a plane, called the ‘ecliptic’, and our rotational axis is tilted to the plane of the ecliptic by 23.5° (though the angle of tilt varies between 21.8° and 24.4° over an approximately 41000-year cycle). This means that from March to September the northern hemisphere is inclined toward the Sun and from September to March the southern hemisphere is inclined inwards, bringing summer to the two hemispheres in turn. Figure 2.17 shows how the tilted axis produces our seasons; these are labelled for the northern hemisphere and the names should be reversed for the southern hemisphere (winter becomes summer,

spring becomes autumn). Only at the spring and autumn equinoxes is the noonday Sun directly overhead at the equator. At noon on 21 June it is directly above the tropic of Cancer and at noon on 21 December it is directly above the tropic of Capricorn. This is how the two tropics are defined.

Our orbit around the Sun is not circular, but slightly elliptical. We are at our closest to the Sun (perihelion) on 3 January and furthest away (aphelion) on 4 July and, therefore, we receive about 7 per cent more solar radiation in January than we do in July. This should make southern-hemisphere summers warmer than those in the northern hemisphere, and the winters cooler, but in reality the situation is reversed. This is due partly to the masking of so small an effect by the general circulation of air, and partly to the fact that at present the northern-hemisphere summer

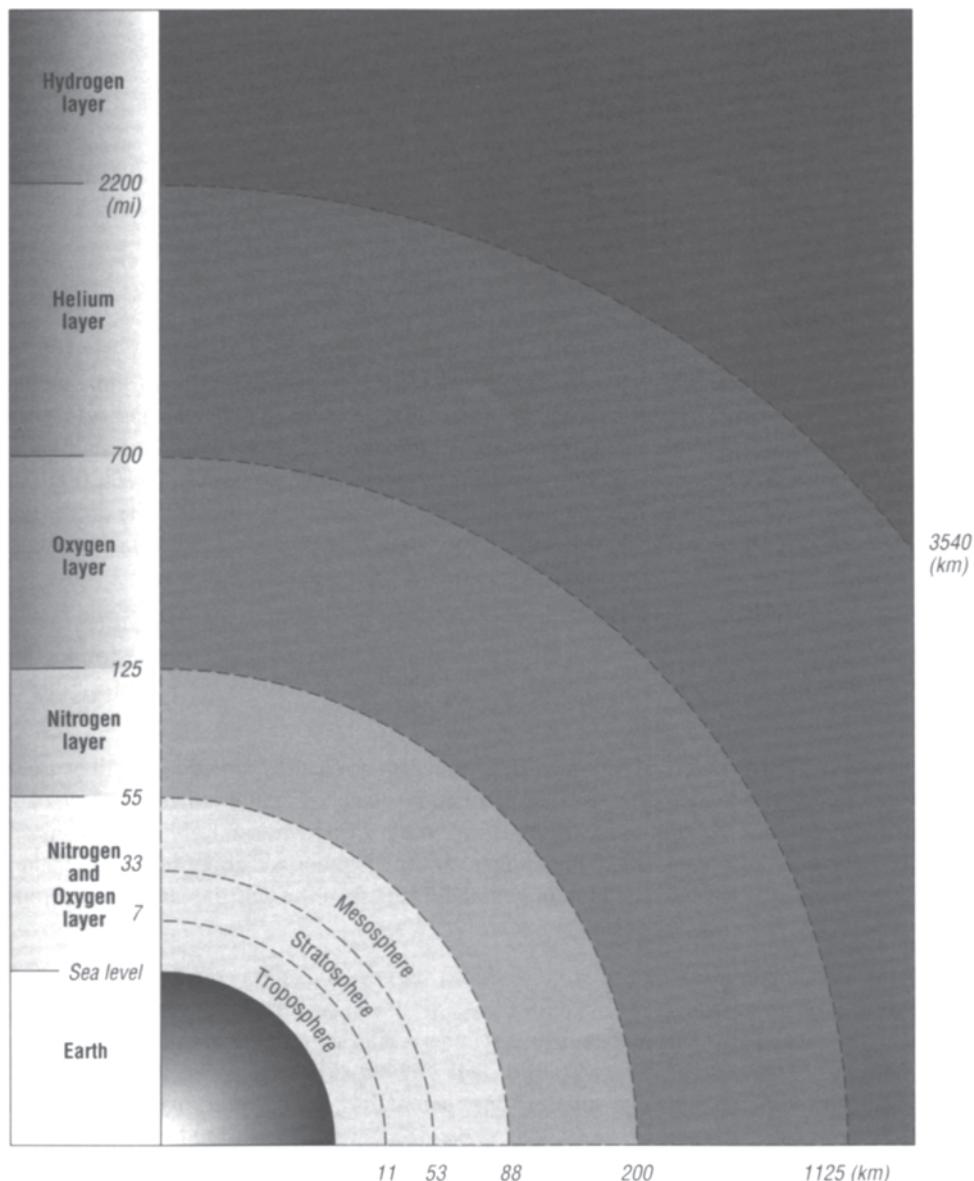


Figure 2.16 Chemical composition of the atmosphere with height

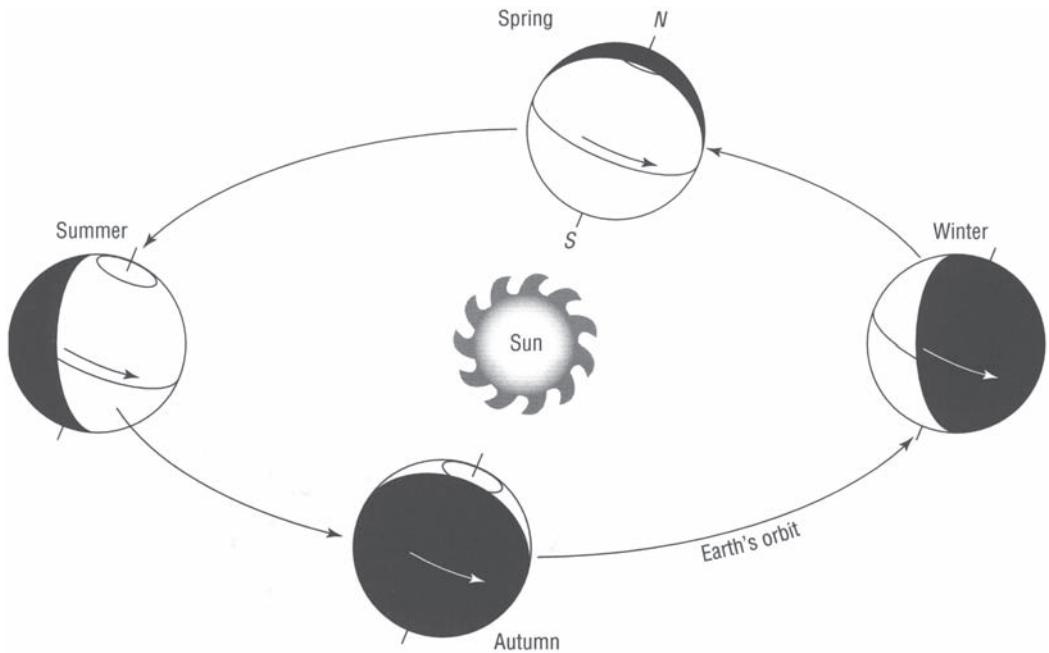


Figure 2.17 Seasons and the Earth's orbit

(March to September) is five days longer than the winter (September to March), a situation that changes slowly. The gravitational attraction of the Sun, Moon and, to a much smaller extent, the planets on the slight bulge around the Earth's equator cause the position of the Earth at the equinoxes to move westward by 50.27" ("=seconds of arc) a year, so it takes 25800 years for them to complete a full cycle and return to their starting position. This is called the precession of the equinoxes and it alters the dates at which the Earth is at perihelion and aphelion. About 13000 years from now we will be at aphelion (furthest from the Sun) in January.

It is within the tropics that the surface receives the most intense insolation and, therefore, is heated most strongly. The movement of air and oceans then carries heat from the tropics into higher latitudes. This transport of heat produces the main 'climates' of the world and our day-to-day weather.

To either side of the equator, the prevailing winds are from an easterly direction and they are so reliable that sailing ships made extensive use of them. The name 'Trade Wind' has nothing to do with commerce. 'Trade' used to mean 'course' and 'to blow trade' meant to blow in a constant direction. Such was their importance that eminent scientists theorized about their cause and it was from their calculations that the first understanding arose of the way the atmosphere transfers heat.

In 1686, the astronomer Edmond Halley (1656–1742) suggested that hot equatorial air rises and is replaced by cooler air from higher latitudes. He was almost correct, but he could not explain why the returning air arrived from the north-east and south-east, rather than from due north and south. This was explained in 1735 by George Hadley (1685–1768). He realized that the Earth rotates beneath the air, changing its apparent direction of flow, but a century passed before the French physicist Gaspard Gustave de Coriolis (1792–1843) discovered what really happens. In 1835 he proposed what is now known as the 'Coriolis effect' (or 'force', although no force, in a mechanical sense, is involved).

On the surface of a rotating sphere, the speed at which any point moves depends on its latitude, because that determines the distance it must travel: a point at the equator travels faster than one at a higher latitude. Air at the equator is moving eastward at the same speed as the surface beneath it. If it moves away from the equator, because it is not attached to the surface, it continues to travel eastward at the same speed (slowing steadily by friction), which is now faster than the surface beneath it, so its motion has an eastward component in relation to the surface. Similarly, air moving towards the equator is moving eastward more slowly than the surface and so it appears to drift westward, in fact because the surface is overtaking it. This is the Coriolis effect and it accounts for the fact that air does not move north or south along straight paths in relation to the surface. The strength of the Coriolis effect increases with distance from the equator and in 1865 the American meteorologist William Ferrel (1817–91) pointed out that in low latitudes the conservation of angular momentum would be more influential than the Coriolis effect (BARRY AND CHORLEY, 1982, p. 138).

What Hadley appreciated was that air, warmed at the equator, will rise, cool as it rises, and descend again. This establishes a convective cell of moving air which drifts to the east as it moves away from the equator and to the west as it returns. We identify winds by the direction from which they blow (i.e. the opposite of the direction in which they blow), so this accounts for the easterly trade winds experienced at the surface in the tropics and also proposes westerly winds at high altitude. This tropical cell, in reality a system of several cells, is known as the ‘Hadley cell’.

Rising equatorial air produces a region of predominantly low surface atmospheric pressure. The air cools near the tropopause and descends in the subtropics, where surface pressure is predominantly high. Very cold, dense air also sinks over the poles, forming polar high-pressure regions. Air flows out from these regions at low level and rises again in middle latitudes, forming a second set of cells. These two drive a third, mid-latitude system of cells, comprising air flowing away from the equator at low level, rising where it meets air flowing in the opposite direction from the pole, and dividing so that some of its air feeds the polar cell and some returns equatorward, descending in the subtropics with the descending air from the Hadley cell. Figure 2.18 illustrates this flow and the winds associated with it. The intertropical convergence zone (ITCZ) is where the two tradewind systems meet. Air rises gently, producing a low-pressure belt near the surface, and this in turn often results in calm air, the region sailors called the ‘doldrums’, a name it acquired in the middle of the last century.

Where the cells meet, near the tropopause in the subtropics and again at about 60°, there is a sharp difference in temperature in the air to either side of a boundary. These temperature gradients produce strong westerly air flows, known as the ‘jet streams’. The subtropical jet stream is fairly constant. The polar front jet stream, at an altitude of 9–15 km, is much more irregular, varying in its latitudinal location and sometimes disappearing altogether, but it also produces the strongest winds. These can reach 150–250 km h⁻¹ and more than 450 km h⁻¹ in winter, when the temperature gradient is strongest.

Atmospheric convection cells produce the boundaries (fronts) between bodies of air at markedly different temperatures which give rise to the jet streams, but it is the jet streams that dominate the weather conditions experienced below. This is especially true of the polar front jet stream, which develops waves and then breaks into cells. This process most commonly occurs in February and March in the northern hemisphere, each complete cycle lasting several weeks. Figure 2.19 illustrates four stages in what is known as the ‘index cycle’. At first (1), the winds are zonal (i.e. they flow steadily from west to east) and there is little mixing of the air to either side. Waves start to develop (2) as the jet stream widens and its velocity increases. Air is now flowing more to the north and south and so it is influenced by the Coriolis effect and the conservation of angular

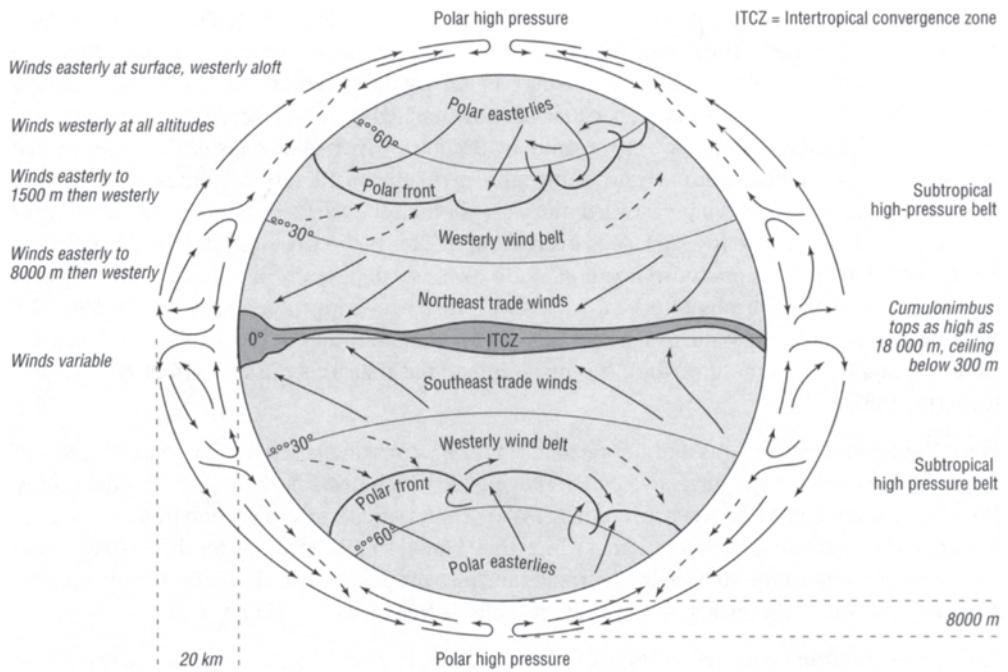


Figure 2.18 General circulation of the atmosphere

Note: ITCZ=intertropical convergence zone

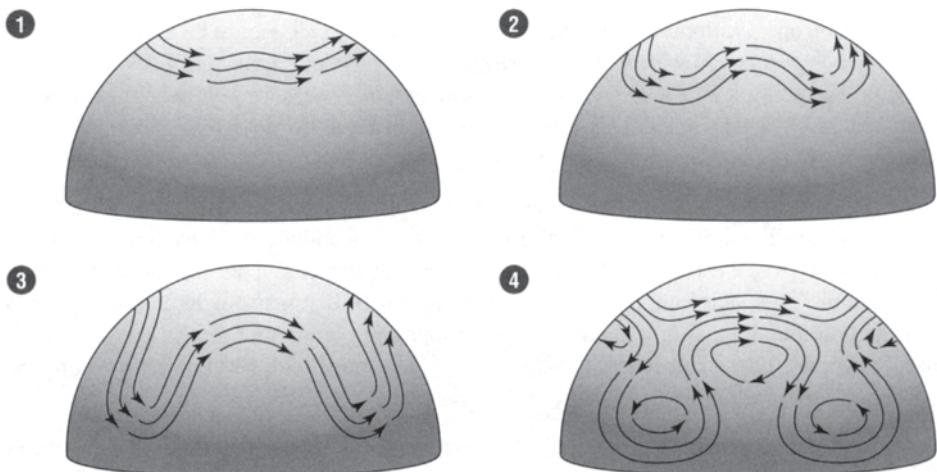


Figure 2.19 The development of cells in jet streams and high-level westerlies

momentum. These tend to exaggerate the undulations, which become extreme (3). Finally (4), the wind pattern fragments into cells. After this, the jet stream disappears for a short time, then re-establishes itself and the index cycle repeats. During 1, surface pressure systems and the weather associated with them move steadily in an easterly direction, but during 2 and 3 the movement becomes more irregular, with latitudinal deviations. In 4, the weather becomes almost stationary,

with mild high-pressure areas in high latitudes and cold low-pressure areas in lower latitudes blocking the easterly movement.

The unsettled nature of the climate at latitudes around 50–60° is due to the proximity of the polar front and the vagaries of the jet stream associated with it: surface conditions are being determined by events at a height of some 10 km. As the front moves north and south, places in about the latitude of Britain may be exposed alternately to polar and tropical air masses and to pressure systems that move rapidly or remain stationary for weeks at a time. Should the global climate change, becoming generally warmer or cooler, there could be a more permanent shift in the location of the polar front, with quite profound implications for places in this meteorologically critical region.

Weather comprises more than temperature, of course, and temperature is dependent on more than the simple convective transfer of heat. Both air temperature and surface weather involve the evaporation, condensation, and transport of water by large masses of air that have acquired distinct characteristics over the oceans or continental interiors then moved into different areas, and on reactions between adjacent air masses with different properties.

The convection-cell model provides only a general, rather crude account of the general circulation of the atmosphere, by which heat is transported from low to high latitudes. It takes no account of the transport of heat by the oceans, for example. This is of major importance and, as the El Niño and NADW-Dryas episodes demonstrate (see the next section), apparently minor disturbances can produce dramatically different situations. The atmospheric system is nothing if not dynamic!

16 Oceans, gyres, currents

In the early 1990s, southern Africa experienced its worst drought this century. Nearly 100 million people went short of food. This was an extreme example of a change that brings abnormal weather to many parts of the world every few years, associated with a rise in the sea-surface temperature off the north-western coast of South America. That an apparently minor warming of the sea, of not much more than 3°C, can have so profound an effect half-way round the world demonstrates the degree to which our climate is influenced by the oceans. The link between the temperature change and climate is now so well established that the surface temperature in the eastern Pacific just south of the equator can be used to predict maize yields in Zimbabwe up to a year ahead, with an even more advanced early warning of what to expect from the ability scientists now have to predict the temperature changes themselves up to a further year ahead (CANE *ET AL.*, 1994).

The circumstances which have such drastic consequences in Africa are called ‘El Niño’ or, to give the phenomenon its full title, an ‘El Niño-Southern Oscillation event’, or ENSO. Figure 2.20 shows the parts of the world most seriously affected, but minor effects are felt even further afield. Northwestern Europe, for example, often has a cool, wet summer following an ENSO. In the areas most seriously affected, temperatures deviate by up to 0.5°C from normal, but deviations of about 0.2°C occur throughout the northern hemisphere. As well as droughts in various parts of the world, a recent ENSO event was blamed for unusually warm weather in Alaska, a remarkably warm winter in the eastern United States, a 100-mm rise in sea level and severe beach erosion in California, the death of coral reefs in the Pacific, and a variety of diseases ranging from bubonic plague to encephalitis in the United States!

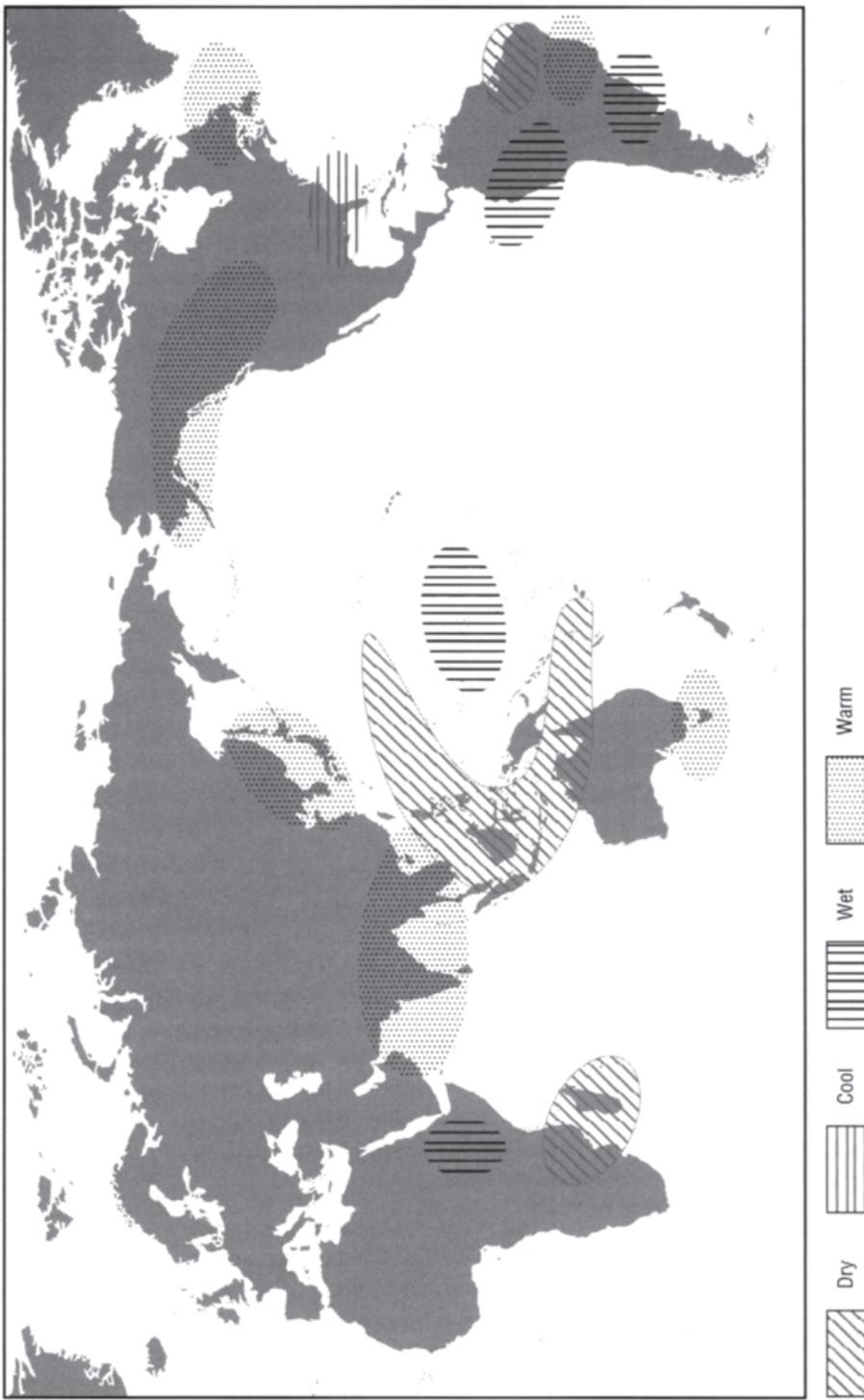


Figure 2.20 Weather changes associated with El Niño-Southern Oscillation events

Ordinarily, the South Equatorial Current, driven by the trade winds, carries surface water westward, away from South America and towards Indonesia. Beneath this fairly shallow layer of surface water there is a boundary, the thermocline, below which the temperature drops sharply. The westward movement of surface water is compensated by an easterly flow along the thermocline, known as the Equatorial Undercurrent, or Cromwell Current. This system moves warm surface water westward, so the surface layer, above the thermocline, is about 200 m deep around Indonesia but very shallow off the American coast, where the thermocline almost reaches the surface. The sea-surface temperature is between 27.5°C and 30°C; this is close to the maximum temperature sea water can reach, because it is cooled by evaporation and the higher its temperature the more water evaporates from it.

The change begins with the distribution of atmospheric pressure over the Indian and Pacific Oceans. This causes the intertropical convergence zone (ITCZ) to move further south than it usually does in the midwinter months of December to February. This is the ‘Southern Oscillation’ and it causes the southern-hemisphere trade winds to weaken or even change direction. The wind driving the South Equatorial Current weakens or reverses, and so warm surface water begins to accumulate off South America, sometimes reinforced by water being driven by the wind from the west. Air moving across the warm water becomes moist and loses its moisture when it reaches the coast. The resulting rains brings abundant grazing for livestock along the arid coastal region, heralding an *año de abundancia*, and because it usually commences around Christmas the phenomenon is known as El Niño, ‘the (boy) child’. In some years the opposite happens. The ITCZ stays far to the north, the trade winds strengthen, and the water off South America is colder than usual. This is known as ‘La Niña’ (HIDORE AND OLIVER, 1993, pp. 169–178).

Although El Niño brings abundance to the farmers of the coastal strip, most other people suffer, especially the fishermen. The Humboldt or Peru Current, flowing northward along the western coast of South America, carries nutrients collected on its long journey from Antarctica, and off the coast of Peru and Ecuador its rich, cold waters well up to the surface through the thin warm-water layer. These upwellings sustain abundant populations of marine plants and animals and support an important anchovy fishery. During an El Niño, however, the nutrient-rich water ceases to reach the surface and the fishery fails.

ENSO events happen roughly every seven years. They are not new: the first was recorded in 1541 and more recently they happened in 1891, 1925, 1953, and 1972–83. There was a particularly strong one in 1982–3, the effects of which were still being detected ten years later. Severe ENSO events also occurred in 1986–87, 1995–96, and 1998–99 (www.elnino.noaa.gov/lanina_new-faq.html). The change in the flow of equatorial water caused long-period waves (Rossby waves) that crossed the North Pacific and shifted the flow of the Kuroshio Current northwards, bringing warmer surface waters to the mid-latitude Pacific that were still being detected in 1993 (JACOBS ET AL., 1994).

Ocean currents play an important part in transporting heat from the equator to high latitudes. The currents themselves are mainly driven by the prevailing winds, but as water moves away from the equator or towards it the motion is influenced by the Coriolis effect. This results in the generally circular currents, or ‘gyres’, of the North and South Pacific, North and South Atlantic, and Indian Oceans. Figure 2.21 shows the principal gyres and currents and their directions of flow.

With centres about 30° north and south of the equator, the gyres flow in a clockwise direction in the northern hemisphere and an anticlockwise direction in the southern hemisphere. Generally, the climatic effect is similar in both hemispheres. Water travelling from the poles towards the equator passes close to the western coasts of the continents, cooling them, and

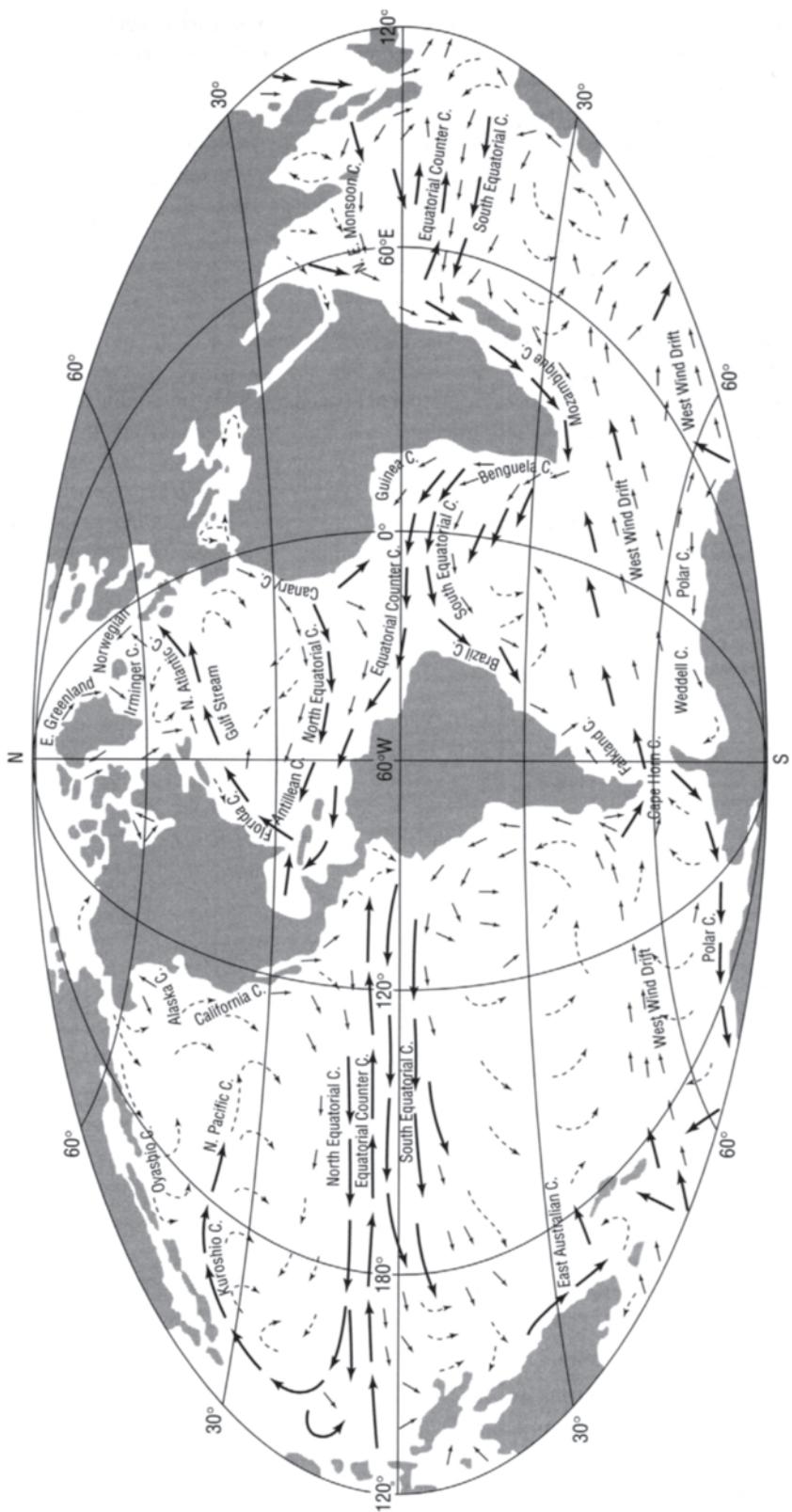


Figure 2.21 Ocean currents

equatorial water flowing poleward passes the eastern coasts, warming them. It is possible to see from this how a northward deflection of the warm Kuroshio Current, off Japan, could warm the northern Pacific with climatic consequences felt in California.

There is one important exception to the general rule and it provides another example of the strong link between ocean currents and climate. In the North Atlantic, the North Equatorial Current flows westward into the Caribbean, then turns north through the Gulf of Mexico, where it becomes first the Florida Current, then the Gulf Stream. The Gulf Stream flows north-eastwards across the Atlantic. It turns south again in the latitude of Spain and Portugal, but a branch continues to flow north-east, passing the shores of Britain. This is the North Atlantic Drift (or Current) and it gives Britain a much milder climate (with palm trees growing in western Scotland and almost subtropical conditions in sheltered places on the Isles of Scilly) than Newfoundland, which in fact is to the south of Britain but cooled by the Labrador Current. Should the Gulf Stream alter its behaviour so that the North Atlantic Drift ceased to break away from it and the entire current turned south together, Britain and north-western Europe would experience a much colder climate.

Atlantic conveyor

Near the edge of the Arctic sea ice, water is close to the temperature at which it is densest and as ice forms the crystallization process separates water molecules from salt molecules. This increases the salinity of adjacent water and, because it contains more salt, the density of the water also increases. This water sinks all the way to the bottom of the Atlantic, flowing south as the North Atlantic Deep Water (NADW). It crosses the equator and continues to the edge of the Antarctic Circle, where it joins the West Wind Drift, or Circumpolar Current, flowing from west to east. From there the water drives systems of currents in all the oceans, eventually returning to the North Atlantic.

The NADW removes cold water from the North Atlantic. This water remains cold until it rises in the South Pacific. During its progress through equatorial and tropical regions the water warms and it returns as warm water to the area near Greenland, where it replaces the NADW that is sinking and moving south.

This circulation is of major importance in regulating the climates of the world.

This is believed to have happened about 11000 years ago and, apparently paradoxically, to have been caused by rapid climatic warming. At that time, the last (Devensian) glaciation was coming to an end and the ice sheets were melting. The Laurentide ice sheet, which covered much of northern North America, discharged its water into the North Atlantic, mainly through the Mississippi River system, releasing a huge amount of very cold fresh water which floated above the denser sea water. This was part of the reason for the failure of the North Atlantic Drift. The other concerned the formation of North Atlantic Deep Water (NADW). At the edge of the sea ice, sea water is very dense. When sea water freezes, its salt is removed as the ice-crystal lattice forms, so adjacent water is more saline than average. At the same time it is also at about 4°C, the temperature at which water is densest. The dense water sinks beneath the less dense water and forms a slow-moving current close to the ocean floor, flowing towards the equator. This is the NADW and the sinking water is replaced by northward-flowing surface water. It is this system that controls the

North Atlantic gyre and the North Atlantic Drift which is part of it. The warming at the end of the glaciation was associated with a retreat of the sea ice as the polar front moved as far north as Iceland, and this disrupted the formation of NADW.

About 11000 years ago and for about 1000 years Western Europe was plunged back into ice-age conditions. When the reversal started, Scotland may have been entirely clear of ice, but before long much of the country lay beneath an ice sheet hundreds of metres thick (ROBERTS, 1989, pp. 50–53). This rapid and dramatic deterioration in climate was first detected by the presence of pollen grains of *Dryas octopetala* (mountain avens) in soils that could be dated to this period. *Dryas octopetala* is a plant characteristic of alpine and subarctic conditions and it has given its name to the climatic reversal, which is known as the Younger Dryas. This lasted from about 11000–10000 years BP (before present) and it is called ‘Younger’ because there is some evidence of an earlier cooling, prior to 12000 years BP, known as the Older Dryas (PENNINGTON, 1974, pp. 32–34), which may have had a similar cause.

Ocean currents are clearly defined and some are fast-flowing. The Kuroshio Current, for example, flows at up to 3 m s^{-1} . Their climatic effect is indirect, however, in that they affect the properties of the air in contact with them, rather than the water warming or cooling the coasts with which it makes contact, and it is the air which brings weather to the continents. The link is obviously of the greatest importance, and any long-term climatic prediction must be based on a much greater understanding of it than exists at present. Is it feasible, for example, that a rapid climatic warming in the northern hemisphere due to the greenhouse effect might disrupt the formation of NADW and trigger a severe cooling in north-west Europe? What causes ENSO events and are they likely to become more or less frequent should there be a global change in climatic conditions? At present such questions cannot be answered. Until they are, all predictions of the regional implications of climate change must be approached with great caution.

17 Weather and climate

Rain, snow, sunshine and showers, wind and storm, are among the phenomena that constitute our weather, the conditions we experience from day to day in a particular place and which weather forecasters aim to predict. The average weather conditions experienced over a large area year by year constitute the climate of that area. The two concepts, of weather and climate, are distinct and provide the subject matter of two equally distinct scientific disciplines: meteorologists study weather and climatologists study climates. Obviously the two overlap, for you cannot understand one without a fairly detailed knowledge of the other. When we discuss the greenhouse effect, for example, we base our ideas on studies made by climatologists; when we wish to know whether it would be wise to plan a picnic for the weekend we consult a meteorologist.

Weather phenomena result from interactions between bodies of air and water at different temperatures and are governed by a small number of general principles. Because air is compressible, atmospheric pressure decreases with height above sea level. If a ‘bubble’ of air (known technically as a ‘parcel’ of air) is made to rise, therefore, its volume will increase as it enters regions where pressure is lower. As it expands, the air becomes less dense. This means its molecules are further apart, and in ‘making more room’ for themselves molecules push one another aside, an activity which requires them to expend energy. Having less energy, the molecules move more slowly and as they slow down, so the parcel of air cools. There is no exchange of heat with the surrounding air; the cooling involves only the expanding parcel. Similarly, if air descends it is compressed, acquires energy, and warms. This cooling and warming is called ‘adiabatic’ and it is a version of the first law of thermodynamics

(which states that in an isolated system, such as a parcel of air, the total internal energy of the molecules can be changed only by work done by them or on them). The concept of total internal energy explains why cold air high above us does not simply sink to the surface: were it to do so it would warm adiabatically to a temperature as high as, or higher than, that of air near sea level; although it is cold, its ‘potential temperature’ is high (HIDORE AND OLIVER, 1993, p. 112).

The rate at which air temperature decreases with height in the troposphere is called the ‘lapse rate’. The average sea-level temperature is 15°C, the average tropopause temperature -59°C, and the average height of the troposphere 11 km, so the ‘standard’ lapse rate is about 6.7°C km⁻¹. The actual lapse rate, called the ‘environmental’ lapse rate, varies from the standard according to local conditions, and if air is cooling (or warming) adiabatically its rate of temperature change depends on the amount of moisture it contains. For dry air, the dry adiabatic lapse rate is 10°C km⁻¹, but if the cooling triggers the condensation of water vapour the cooling air will be warmed by the latent heat of condensation, so the saturated adiabatic lapse rate is lower than the dry adiabatic lapse rate. Its value varies with the amount of condensation, but averages about 6°C km⁻¹.

Adiabatic cooling and warming

When air rises it expands, because there is less weight of air above it in the column reaching to the top of the atmosphere and, therefore, atmospheric pressure decreases with increasing altitude. As a gas expands, its molecules move further apart. In doing so they expend energy and, since the molecules have less energy, the temperature of the expanded air decreases.

It follows that a rising ‘parcel’ of air will cool without exchanging energy with the surrounding air. This is ‘adiabatic’ cooling (from the Greek *adiabatos*, ‘impassable’).

A descending ‘parcel’ of air warms by the same adiabatic process. As it enters a region of higher pressure it is compressed and gains energy, which heats it.

Water vapour is a gas and the amount of it in a particular body of air is expressed as the ‘humidity’ of the air. Humidity can be measured in several ways. Absolute humidity is the mass of water vapour in a given volume of air, and specific humidity is the mass of water vapour in a given mass of air (including the water vapour). Relative humidity, which is the measure most widely used, is the proportion of water vapour in relation to the amount required to saturate the air, and is given as a percentage. Warm air will hold more water vapour than cool air, so relative humidity varies with temperature as well as actual water-vapour content. The converse of this is that if air is cooled, a temperature will be reached at which its water vapour condenses. This is the ‘dewpoint’ temperature.

When water changes phase, between solid and liquid, liquid and gas, or directly (by sublimation) between solid and gas, energy is either absorbed or released, as latent heat. Latent heat warms or cools the surrounding air. It is why the air feels warmer when snow starts to fall and cooler when ice thaws, and it also governs the dynamics of storm clouds, hurricanes, and tornadoes. The amount of heat is considerable. When 1 gram of water evaporates, 2500 joules of energy is absorbed and the same amount is released when water vapour condenses; the change between liquid and solid requires the absorption or release of 334.7 J of energy; and sublimation between solid and vapour requires the absorption or release of the sum of these, 2834.7 J for every gram.

Over the continents and oceans, mixing of the air tends to equalize the pressure, temperature, lapse rate, and humidity horizontally over a large area. Such air is called an ‘air mass’. Its characteristics are determined by the region in which it formed, so air masses can be labelled. The first division is between those formed over continents and those formed over oceans, the second between those formed in arctic, polar, and tropical latitudes. This yields six types: continental arctic (cA), continental polar (cP), continental tropical (cT), maritime arctic (mA), maritime polar (mP), and maritime tropical (mT). Continental air is dry, maritime air moist and, as the names suggest, tropical air is warmer than polar or arctic air; temperatures are more extreme in continental than in maritime air, owing to the moderating influence of exposure to the oceans. Air masses do not remain stationary, of course. They move, and as they do so they pass into new regions which modify their characteristics. A continental air mass originating over North America becomes a maritime air mass by the time it has crossed the Atlantic and reached Europe.

An air mass does not exist in a vacuum. It has borders, beyond which lie other air masses, with different characteristics, and as it moves the air mass ahead of it must also move. Not all air masses move at the same speed, however, so the boundaries are sites where one air mass is displacing another. The theory describing these reactions was developed during the First World War by scientists at the Bergen Geophysical Institute, Norway, led by Vilhelm Firman Koren Bjerknes (1862–1951). Because of the predominant news at the time the theory was being constructed, Bjerknes and his colleagues described the boundary between different air masses as a ‘front’.

Fronts are identified as ‘warm’ or ‘cold’ depending on whether the air behind the front is warmer or cooler than that ahead of it. They meet the ground surface at an angle, the frontal slope, and warm and cold fronts slope in opposite directions (ALLABY, 1992, pp. 86–87). Warm fronts have a gradient of about 1:1000; cold fronts are much steeper, sometimes with a gradient of as much as 1:50, and they generally travel faster than warm fronts. Figure 2.22 illustrates typical conditions in the two types of frontal region. In (1), rapidly advancing cold air moves beneath warmer air, forcing it to rise; as the rising air cools, its water vapour condenses to form cumuliform (heaped) clouds, bringing heavy showers and possibly thunderstorms. In (2), a gently sloping warm front is advancing more slowly, so its air is lifted gently above the cooler, denser air and stratiform (sheet-like) clouds form, bringing drizzle or steady rain.

Cool air is denser than warm air, so its pressure is higher. Air flows from areas of high to low pressure, but because of the Coriolis effect and vorticity (the tendency of a flowing fluid to follow a spiral path) it follows an approximately circular path with a speed proportional to the pressure gradient (the pressure difference between the high and low areas). In the northern hemisphere air flows anticlockwise (cyclonically) around areas of low pressure and clockwise (anticyclonically) around areas of high pressure; in the southern hemisphere these directions are reversed. The difference in speed between cool and warm air masses results from the way air moves around them: in the northern hemisphere air flowing anticyclonically moves faster than air flowing cyclonically and in the southern hemisphere the reverse applies.

At first, a front is simply a line, with cold air on one side, warm air on the other, and air flows in opposite directions on either side. Some air crosses the front, however, and a V-shaped wave appears in the front. As this grows more pronounced, the cold air is partly behind the warm air and overtaking it. A low-pressure centre develops at the apex of the wave; this is a ‘cyclone’ or ‘depression’. The undercutting cold air then starts to lift the warm air clear of the ground; at this stage the fronts are becoming occluded. Occlusion continues until all the warm air has been lifted above the ground; after that the depression dissolves and disappears. Figure 2.23 shows the way clouds are distributed around the depression as it develops and dissipates.

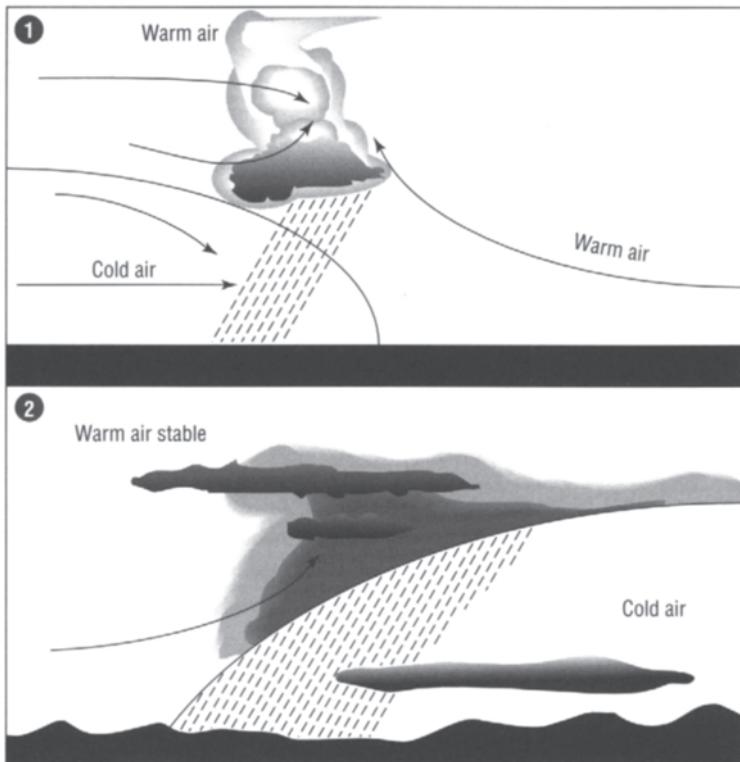


Figure 2.22 *Formation of cloud at a front. 1, Cold air is advancing rapidly, undercutting warm air and forcing it to rise. 2, Stable warm air is advancing, rising gently over cold air*

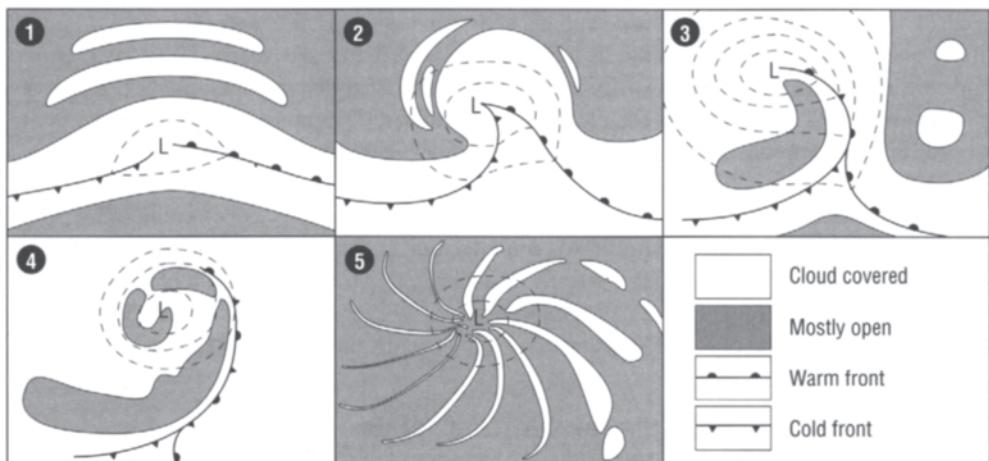


Figure 2.23 *Distribution of cloud around frontal systems. 1, The system starts to form. 2, The two fronts are still separate. 3, Warm air is rising over cold air. 4, The fronts are almost fully occluded. 5, Cloud distribution around a low-pressure centre*

Frontal systems of this general type, interspersed with periods of settled weather as the cold and warm sectors pass, typify mid-latitude climates. They can also bring severe weather, especially in the warm sector ahead of a cold front. Where warm surface air is lofted by an advancing cold front and winds at high altitude flow from a different direction, a moving line of thunderstorms, called a squall line, may occur.

Water vapour, condensing in the rising air, releases heat, so the air continues to rise and water vapour is constantly evaporating and condensing. In a fair-weather thunderstorm, caused by locally strong heating of the ground, the upward development is limited by precipitation falling from the top of the cloud, which slows the updraughts, but in a squall line the high-level wind shifts the top of the cloud to one side. This removes the limiting factor, because the precipitation falls to the side of the cloud rather than through its centre. The updraughts carry water droplets, which freeze near the cloud top, partly melt as they fall, then freeze as they rise again, growing the ‘onion-skin’ layers of ice that make hailstones until they are heavy enough to fall from the cloud, sometimes melting again in the warmer air between cloud and ground.

The vertical motion leads to a separation of electrical charge. Positively charged ice crystals accumulate near the top of the cloud, negatively charged particles near the bottom, and a positive charge is induced in the ground beneath the cloud. Lightning discharges these differences as soon as they are big enough to overcome the electrical resistance of the air, the energy released by the lightning causing an explosion of the air adjacent to the flash, which we hear as thunder.

There are many variations to this general scheme. It is not only a cold front that may force warm air to rise, for example. Air also rises as it crosses high ground and may lose moisture as a consequence. The western side of Britain has a wetter climate than the eastern side, because air from the Atlantic loses moisture as it crosses the hills, especially the Pennines and the mountains of Wales and Scotland. Air forced over a mountain may cool adiabatically to below the temperature of surrounding air. On the lee side of the mountain it descends again, warming adiabatically as it does so, to form a warm, dry wind, called a Föhn wind in Europe and a Chinook in North America. Cold air may also flow down a hillside, to produce ‘frost hollows’. Valleys may funnel the wind, intensifying it. The cold mistral of southern Europe is caused by funnelling along the Rhône valley.

Such local and regional effects mark the difference between the weather people living in a particular place experience and the climate, which is the average of the weather conditions over an entire region. Aided by satellite observations of weather systems and their direction and rate of movement, a wide network of surface reporting stations on land and at sea, and immense computing power, it is possible nowadays to forecast weather with a fair degree of accuracy for a few days ahead. Weather systems are extremely complex, however. Two meteorological situations that appear identical can develop in radically different ways over a matter of days or weeks. This is because although they appeared identical, in fact they differed in ways too small to be measured and these differences directed their development. This extreme sensitivity to minute variations is characteristic of ‘chaotic’ behaviour, the most famous analogy being ‘the butterfly effect’, in which the flapping of its wings by a butterfly in Beijing alters the development of storm systems a month later in New York. This makes it impossible to forecast weather further ahead than a few days; indeed, such long-range forecasting may be inherently impossible (GLEICK, 1988, pp. 9–31).

18 Glacials, interglacials, and interstadials

Jean Louis Rodolphe Agassiz (1807–73) was a zoologist with a particular interest in fishes, both living and extinct. He was appointed professor of natural history at the University of Neuchâtel in

his native Switzerland and there he might have remained, the most eminent ichthyologist of his generation, were it not for what began as a secondary interest.

Switzerland is noted for its glaciers and the Swiss are very familiar with them. In various parts of Europe there are boulders and gravels made of rock quite different from that of the region in which they are found, and in the 1830s some Swiss scientists were speculating that these rocks had been pushed into their present positions by glaciers. If so, it implied first that glaciers move and, second, that they once extended much further than they do now. Agassiz was sceptical, but decided to test the idea and in 1836 and 1837 he spent his summer holidays studying Swiss glaciers.

He soon discovered piles of rock to either side of glaciers and where they terminated, and noticed that some of these rocks were scoured with lines, as though small stones had been dragged across them under great pressure. Continuing with his visits to glaciers, in 1839 he found a hut, built on a glacier in 1827, a mile from its original location. His final test involved fixing a line of stakes across a glacier from side to side. Two years later, in 1841, he found they had moved and now formed a U shape, because the stakes near the centre of the glacier had moved further than those near the sides.

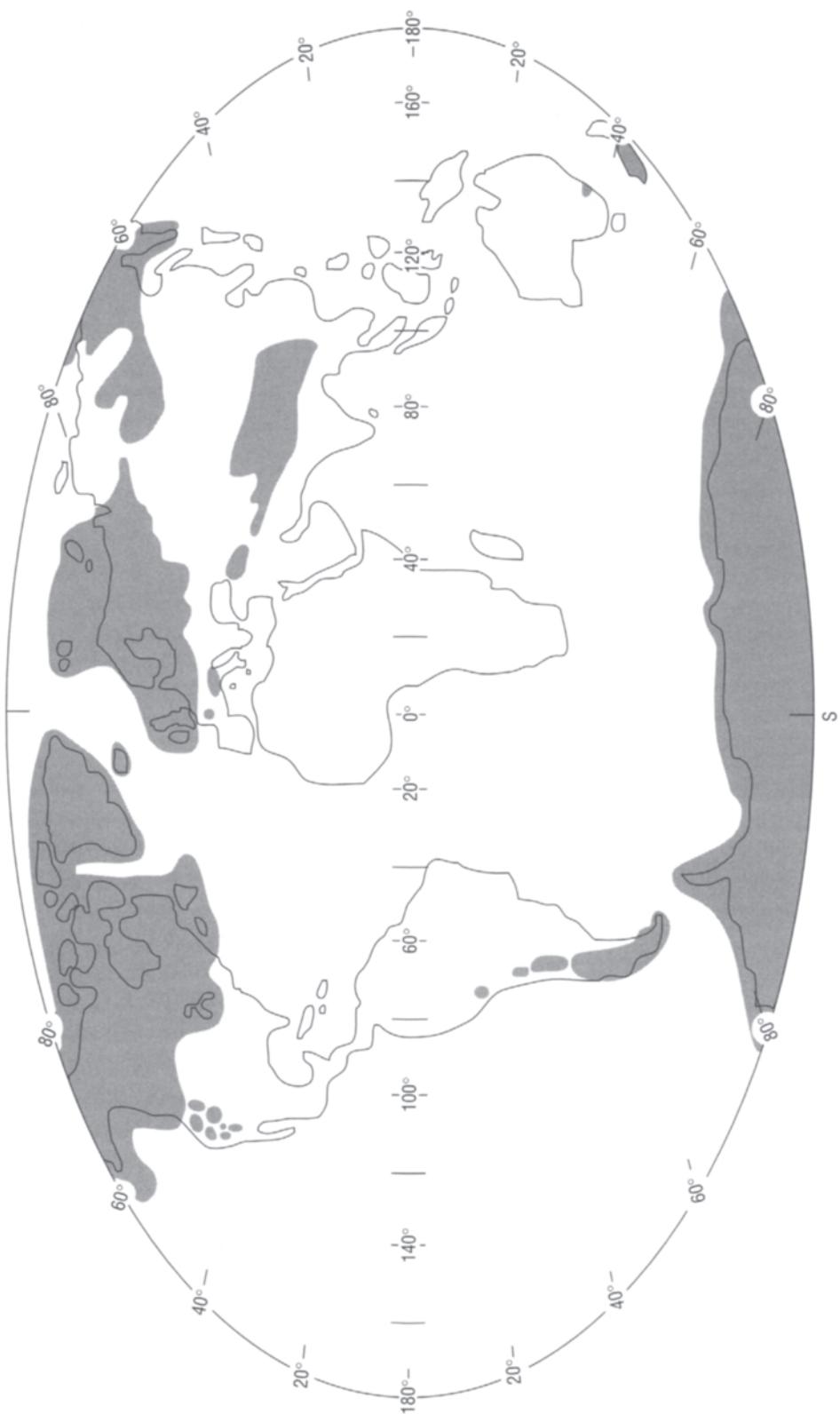
Already persuaded that glaciers do move, Agassiz published his ideas in 1840, as *Études sur les glacières*, shortly before his rival, Jean de Charpentier (1786–1855), published his own version of the same theory. This proposed that in the fairly recent past all Switzerland and all those regions of Europe in which unstratified gravel occurs had been covered by sheets of ice similar to those that still cover Greenland. Agassiz extended his studies to other parts of northern Europe, and concluded that the ‘Great Ice Age’ had been very extensive. In 1846 he was invited to lecture in the United States, mainly because of his work with fossil fish, but he used the opportunity to deliver popular lectures about the Great Ice Age and to seek, and find, evidence for glaciation in North America. He remained in the United States, spending most of his time at Harvard, and became an American citizen.

Boulders and unstratified gravel, unrelated to the underlying rocks, have clearly been transported. Agassiz supplied an explanation for the mechanism of their transport, but an alternative explanation already existed. Many scientists believed the Earth had once lain beneath water, perhaps the biblical flood. In overturning this conjecture, Agassiz had a profound influence on our ideas of Earth history.

The Great Ice Age was soon accepted, but in modern times the original concept has been greatly modified. Figure 2.24 shows the extent to which ice sheets have covered the Earth at one time or another during the last 2 million years. As the map indicates, recent ice ages have affected both hemispheres, although glaciation has been more marked in the northern hemisphere, where land extends into higher latitudes. Much of what appears as open sea between ice sheets was in fact frozen, as the sea around the North Pole is frozen today.

Conventionally, the ice ages are said to occupy the Pleistocene Epoch, which began about 2 million years ago and ended about 10000 years ago, with the end of the last glaciation and the commencement of the Holocene (or Recent) Epoch, in which we now live. In fact, glaciation began somewhat earlier, rather more than 3 million years ago, as glaciers advanced and then retreated again several times (GENTRY AND SUTCLIFFE, 1981), and it is possible that the date of the beginning of the Pleistocene may be revised. Indeed, it may be that our use of the word ‘Holocene’, or ‘Recent’, is premature. Most palaeoclimatologists (scientists who study the climates of the distant past) agree that we are presently living in an interglacial, called the Flandrian, and that one day (no one can say when) this will end and there will be another ice age. If this is so, perhaps the Pleistocene Epoch has not yet ended!

Figure 2.24 Parts of the Earth covered by ice at some time during the past 2 million years



Although we usually think of the Pleistocene as the ‘Great Ice Age’, this is misleading. There was not a single ice age (or glaciation; the terms are synonymous) but several, and in the interglacials separating them climates were sometimes markedly warmer than those of today. During the Ipswichian (or Trafalgar Square) Interglacial, for example, the average summer temperature in southern Britain was some 2–3°C warmer than today and remains of elephants, hippopotami, and rhinoceros, dated to that time (100000–70000 years BP), have been found in what is now central London (hence the name ‘Trafalgar Square Interglacial’). Nor is the Pleistocene the only geological period during which glaciations are known to have occurred. In various parts of the world there may have been glaciations some 2.3 billion years ago and between 950 and 615 million years ago, during Precambrian time, at the end of the Ordovician Period about 440 million years ago, and in the southern hemisphere around 286 million years ago, at the end of the Carboniferous and beginning of the Permian Periods. It does seem, however, that the Earth enjoyed an ice-free period between the Permo-Carboniferous glaciation and the onset of the Pleistocene.

During the Pleistocene, there are now believed to have been four glacial episodes and three interglacials in North America and five glacial and four interglacial episodes in Europe. The present epoch, the Holocene, is also known as the Flandrian Interglacial. This makes it the fifth European interglacial, and acceptance of the name implies that the Pleistocene glaciations have not yet ended; we are still living in the Pleistocene and one day the glaciers will start to grow again. If and when this happens it may do so rapidly, for glaciations can begin quickly and end even more quickly.

Even during glaciations there are periods of remission. Briefer and cooler than interglacials, these are known as ‘interstadials’ (or ‘interstadials’) and are identified by the presence of pollen from plants known to require mild climates. During interglacials, too, temperatures fluctuate. Figure 2.25 shows how average temperatures have varied from 18000 years ago, when the most recent glaciation was at its most severe, to the present day. During the climatic optimum, about 5500 years ago, mid-latitude temperatures were about 2.5°C warmer than those of today; this was the period during which civilizations flourished in Asia Minor. A smaller optimum, peaking around AD 1000, allowed the Vikings to colonize Greenland. Between about 1450 and 1880 temperatures fell during the ‘Little Ice Age’. This is when fairs were held on the frozen Thames and it may be a minimum from which our climate is still warming. As the graph indicates, the weather has been markedly warmer than it is today during several episodes in the past.

Confusingly, glaciations and interglacials have different names in different places and it can be difficult to match them (ALLABY, 1992a, pp. 145–148). The most recent glaciation is known as the Devensian in Britain, the Weichselian in northern Europe, the Würm in the Alps, and

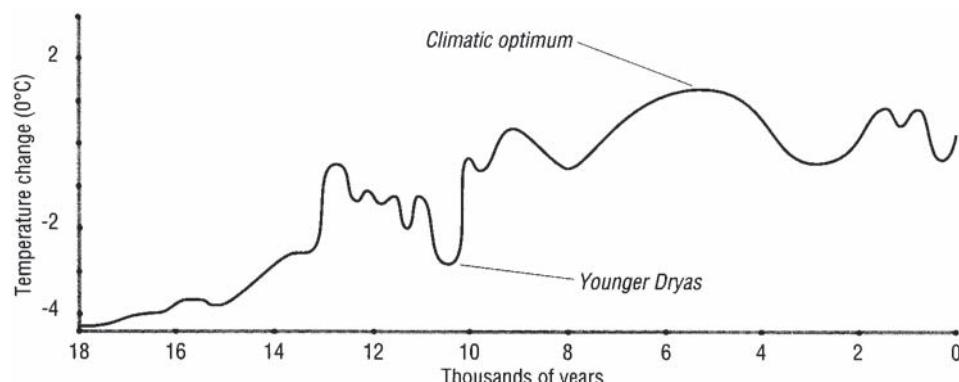


Figure 2.25 Temperature changes since the last glacial maximum (present temperature=0)

is roughly equivalent to the North American Wisconsinian. It began about 70000 years ago and ended about 10000 years ago.

Glaciations and interglacials are associated with climatic changes more extensive than the advance or retreat of ice sheets. Precipitation patterns also change. Today, central Antarctica is possibly the most arid place on Earth and the Arctic is also dry, though less extremely so. Ice sheets accumulate because fallen snow is removed only slowly by ablation, not because precipitation is heavy. At the peak of the Devensian (Wisconsinian) glaciation, 18000 years ago, most deserts were larger than they are today, but during the warm episode following the end of the Younger Dryas the Sahara was smaller than it is now and lakes in that part of Africa were larger (GENTRY AND SUTCLIFFE, 1981, p. 239). Generally, climatic cooling implies increasing aridity, climatic warming increasing precipitation. Indeed, the correlation with rainfall makes it possible to equate interglacials with pluvials (periods with wet climates) and glacials with interpluvials (HOLMES, 1965, pp. 715–716).

Sea levels also change with the advance and retreat of ice sheets. As more and more water reaches the frozen regions and is trapped there, sea levels fall. At times they have been 100 m below their present limit. At times during the Devensian (Wisconsinian), the North Sea was dry land, crossed by rivers, Alaska was joined to Siberia by a wide strip of land, and New Guinea was linked to Australia. The exposure of land bridges presents living organisms with opportunities to expand their ranges and it was during the Pleistocene that humans reached most of the regions in which they live now. About 60000 years ago, people arrived in Australia (MORELL, 1995), presumably having crossed about 100 km of open sea to migrate from Asia to New Guinea-Australia, and some time later there was the first of several migrations from Asia into North America. It is only the more isolated Pacific islands, Madagascar, and New Zealand that have been colonized during the Holocene, about 3000, 2000, and 1000 years ago respectively (ROBERTS, 1989, p. 56).

People did not live on the ice sheets, of course, but the tundra landscapes bordering the ice supported abundant game and the seas provided fish and seals as well as marine invertebrates. Climate change, especially if it is rapid, might be supposed to challenge species and, indeed, many animals became extinct during the Pleistocene. These extinctions are unlikely to have been linked solely to changing climate, however. It is much more likely that they were caused by overhunting as humans expanded their range. Vast quantities of animal bones have been found in certain places, sometimes at the foot of cliffs over which entire herds appear to have been driven, presumably by hunters who then took the meat and other materials they needed from the pile of carcasses. One such site, at Solutré, France, has the bones of more than 100000 horses (ROBERTS, 1989, p. 59), and in North America many large mammals became extinct within 1000 years of the emergence of the earliest human culture. The extinctions were confined to the late Pleistocene and are not linked to earlier climate changes, some of which were at least as rapid; they principally affected large mammals, and they occurred on all continents. Australia lost some 81 per cent of its large mammals 26000–15000 years ago, South America 80 per cent 13000–8000 years ago, North America 73 per cent 14000–10000 years ago, Europe 39 per cent 14000–9000 years ago, and Africa 14 per cent 12000–9500 years ago.

Glaciers scour away soil and their retreat leaves a barren landscape, but one to which plants and animals soon start to return. Holocene recolonization has been well documented, together with the limits species reached before rising sea levels prevented their further migration. The mole (*Talpa europea*), common shrew (*Sorex araneus*), beaver (*Castor fiber*), aurochs (*Bos primigenius*), European elk or North American moose (*Alces alces*), and roe deer (*Capreolus capreolus*) never reached Ireland, for example, because of the formation of the Irish Sea 9200 years ago (SIMMONS ET AL., 1981, pp. 86–89).

Land that is now free from permanent ice was glaciated for rather more than half of the Pleistocene. It is reasonable to suppose that the global climate now fluctuates between glacial and interglacial conditions and that the present interglacial may be near its end, unless the subsequent cooling is overridden by warming induced by the greenhouse effect. The alternation between glacial and interglacial is the most extreme climate change imaginable, but the historical evidence suggests that living organisms adapt to it fairly robustly. As conditions deteriorate enough of them migrate to more favourable environments for their species to survive, from which they return when opportunity allows. In certain local areas, called refugia, communities survive even without migrating, because the older conditions prevail. There are several Pleistocene refugia in Britain, Upper Teesdale being probably the best-known example. Although undoubtedly inconvenient for humans, rapid climate change does not necessarily imply the extinction of species, merely their absence until the next change allows them to return.

19 Dating methods

Consider the world as it might appear to an intelligent mayfly. As it emerges from the stream in which it has spent most of its life, the insect sees a world where the Sun shines, trees are in full leaf, a summer world. To the mayfly, this is how the world is, the only appearance it can ever present. Long before leaves start to fall and even longer before water starts to freeze and the ground is covered with snow, the mayfly will have died.

Unlike the mayfly, we know the world changes, that there is winter as well as summer. Yet our lives, too, are brief and deny us any opportunity to observe at first hand the fact that those aspects of the world we regard as permanent are no more so than the summer sunshine and leaves. Our Earth is changing constantly. Continents move, mountains are thrust upward and then eroded into plains, ice ages come and go, species evolve only to vanish again, but these changes occur on a time-scale that seems long to us. Compared with the 4.6 billion years during which our planet has existed, a human lifespan is ephemeral indeed.

If we are to understand the environment in which we find ourselves, to observe ways in which it may be changing, and to predict future changes and our own influence upon them, we must learn to appreciate the time-scale on which such events occur. We must try to discover how the environment arrived at its present condition if we are to discern trends and compile forecasts. We must study the history of our planet and the first requirement for any historical reconstruction is a reliable means for dating events. We must know when and in what order past events happened.

The first step in reconstructing the past is fairly straightforward. Sedimentary rocks begin as sediments, most deposited beneath water, and they can be seen to form layers. Obviously, the older layers must have been precipitated before those lying above them and changes in the composition of layers must reflect changes in the depositional environment. Unfortunately, sediments seldom remain undisturbed, so although it is easy enough to recognize their layers it is more difficult to determine their relative ages; to do that it is necessary to determine which way up they were when they were precipitated.

It was Georges Cuvier (1769–1832) who first realized that fossils might be used to identify sedimentary strata. He and Alexandre Brongniart (1770–1847), an engineer, applied this idea to a study of the Paris Basin, describing their discoveries in *Descriptions géologiques des environs de Paris*, published in 1811 (BOWLER, 1992, pp. 213–220). Their scheme was based on the observed fact that fossil invertebrates found in some strata are absent from others and, therefore, that certain

Table 2.4 **Geologic time-scale**

Eon	Era	Sub-era	Period	Epoch	Began (Ma)
Priscoan					4600
Archaean					4000
Proterozoic					2500
Phanerozoic					
	Palaeozoic				
		Lower Palaeozoic			
			Cambrian		570
			Ordovician		510
			Silurian		439
		Upper Palaeozoic			
			Devonian		408.5
			Carboniferous		
			Mississippian		362.5
			Pennsylvanian		322.8
			Permian		290
	Mesozoic				
			Triassic		245
			Jurassic		208
			Cretaceous		145.6
	Cenozoic				
		Tertiary			
			Palaeogene		
				Palaeocene	65
				Eocene	56.5
				Oligocene	35.4
			Neogene		
				Miocene	23.3
				Pliocene	5.2
		Quaternary			
			Pleistogene		
				Pleistocene	1.64
				Holocene	0.01

fossil assemblages can be used to identify particular strata wherever those strata are found. In other words, animal species have come into existence, lived for a while, and then have disappeared and their places have been taken by others, newly arrived: we would say evolved. From this it is possible to construct a ‘stratigraphic column’, a vertical section through sedimentary rocks in which each stratum is shown in chronological order. The Cuvier and Brogniart study inspired geologists all over Europe to apply the method to their own localities and eventually to divide geologic time into distinct episodes on the basis of the animals associated with them. Though much amended, the geologic time-scale used today is derived from this work, as are most of the names.

Geologic time is divided into eons, eras, suberas, periods, and epochs. Table 2.4 shows the present arrangement, starting with the oldest, although some of the dates are revised from time to time (Ma means millions of years ago). Priscoan, Archaean, and Proterozoic together comprise the eon formerly known as the Precambrian. The term ‘Precambrian’ is still widely used, but not in a formal sense (all it means, after all, is ‘before the Cambrian’).

The order in which historical episodes should be arranged having been established, the next step is to allot dates to them. The thickness of strata is no help with this. Sedimentation is an

irregular process, so a thick layer may have accumulated rapidly, a thin one more slowly, and there is no way to tell. Some sediments, however, build more regularly, and it was the record they left that allowed the retreat of Scandinavian ice sheets, starting about 10000 years ago, to be traced. Each spring, as the ice melts, an assortment of mineral particles is washed into a lake by the meltwater. Heavier particles, such as sand grains, settle quickly. Later in the year, as water freezes again, the supply to the lake ceases and the finer particles, of silt and clay, gradually settle on top of the sandy layer. Year after year the process is repeated, each pair of layers, one pale and coarse, one dark and fine, being known as a ‘varve’. These can be counted, each varve representing one year, and if varves are forming at the edge of a retreating glacier they will follow it, so that its progress can be traced and dated. The study of varves is known as varve analysis, varve chronology, or a varve count.

Varves resemble tree rings, which provide another method of measuring time. In spring, woody plants grow rapidly by producing large, thin-walled cells in the xylem, just below the bark of stems and branches. Growth slows in summer, ceasing in late summer, and consists of smaller cells with thicker walls. The large cells of spring are pale in colour, the smaller ones of summer dark, and so each year the plant produces a ring of pale wood separated by a thin, dark ring from the pale wood of the following year. A count of the rings is a count of years, but there are some risks. If conditions are very severe, a plant may produce no growth for a whole year, and if conditions are unusually favourable it may produce two or more sets of rings. For this reason, tree-ring dating (called dendrochronology) must be based on as many specimens as is practical, obtained from widely scattered locations. The fact that rings are strongly affected by growing conditions has advantages. The width of rings can be used to infer weather (dendroclimatology) and environmental (dendroecology) conditions at the time they formed.

Obviously, the study of tree rings can provide dates only up to the age of the living plant from which they are taken, but trees can live a surprisingly long time. There are bristlecone pines (*Pinus longaeva*), found in California, more than 4600 years old, and correlating rings from them (taken as cores, without destroying the tree) with rings from dead pines has allowed scientists to construct a chronology for arid zones going back 8600 years and, at the upper tree limit on mountains, one going back 5500 years.

These chronologies are used to calibrate radiocarbon (^{14}C) dates. Bombardment by cosmic radiation generates neutrons, a few of which collide with atoms of nitrogen (^{14}N), displacing a proton and converting the ^{14}N to ^{14}C . Chemically, ^{14}C behaves just like ordinary ^{12}C and living organisms exchange both with their surroundings. When they die, however, carbon exchange ceases. Carbon-14 is radioactive, half of any amount of it decaying to ^{12}C in 5730 ± 30 years (its half-life), so the ratio of $^{12}\text{C} : ^{14}\text{C}$ in dead organic matter is directly related to the time that has elapsed since it died. Radiocarbon dating rests, however, on the assumption that the rate of ^{14}C formation in the atmosphere is constant. This is now known not to be so, because the intensity of cosmic-ray bombardment is variable, but, when correlated with tree-ring series from bristlecone pine, radiocarbon analysis makes it possible to date material up to about 70000 years old.

Dating material older than this requires other methods. These, too, are based on the decay of radioactive elements, but ones with much longer half-lives. The first to be exploited were uranium (U) and thorium (Th). Uranium occurs naturally as a mixture of two isotopes, ^{238}U and ^{235}U in the constant proportions 137.7:1; both decay to stable isotopes of lead (Pb). Uranium-238, with a half-life of 4510 million years, decays to ^{206}Pb , and ^{235}U , with a half-life of 713 million years, to ^{207}Pb . Thorium-232, with a half-life of 13900 million years, decays to ^{208}Pb . Lead also occurs naturally as the stable isotope ^{204}Pb , so this must be deducted from lead isotopes resulting from radioactive decay before an age can be calculated.

Potassium-40, a radioactive isotope of potassium with a half-life of 1300 million years, also occurs naturally (and, because of its presence in our food, is the principal source of our own exposure to radiation). Most ^{40}K decays to ^{40}Ca , which cannot be used because calcium is so common, but about 11 per cent decays by a different route to ^{40}Ar (argon). This decay is used to date rocks more than 250000 years old.

A radioactive isotope of rubidium, ^{87}Rb , decays in a single step to strontium (^{87}Sr) and this decay is used to date certain rocks, especially those containing mica and potassium, but there is some doubt about the half-life of ^{87}Rb . Two values are used: 4.88×10^{10} and 5.0×10^{10} years. A more recent method uses the decay of samarium (^{147}Sm) to neodymium (^{143}Nd). Samarium-147 has a half-life of 2.5×10^{11} years and this decay is used in studies of the formation of rocks in the Earth's crust and mantle (and can also be used on materials of extraterrestrial origin).

It is impossible to predict when an individual unstable atom will decay, but it is possible to calculate the probability that the atom will decay within a certain period. This is called the 'decay constant' for the isotope, from which the half-life can be calculated as the time taken for the decay of half the unstable atoms present. The process is exponential: half the atoms decay in the first half-life period, half the remainder in the second period, half of that remainder in the third, and so on (e.g. 100; 50; 25; 12.5, etc.). Most of the decays used are based on half-lives much longer than the age of the Earth, but it is not necessary to wait until a complete half-life has elapsed before calculating an age. What matters is the ratio of isotopes.

Since radioactive decay involves only the nucleus of the atom, its rate is not affected by temperature, pressure, or any other outside influence. This makes it a very reliable measure of the age of materials. Radiometric dating has allowed scientists to reconstruct the history of the Earth in some detail.

20 Climate change

Milutin Milankovich (1879–1958) spent most of his career as a mathematician and physicist working at the University of Belgrade, where he devoted thirty years to comparing the amount of solar radiation received in different latitudes over the last 650000 years with the climates during that time. He discovered a clear relationship between solar variability and the incidence of ice ages that is now accepted by most climatologists. Presented as a graph, it is known as the Milankovich solar radiation curve (geography.miningco.com/library/weekly/aa121498.htm).

We picture the Earth spinning on its axis and orbiting the Sun in a very regular fashion. So it does, but within its regularity there are slow, cyclical variations. Milankovich identified three that affect climate when they coincide to maximize or minimize insolation.

The first cycle, illustrated in Figure 2.26, concerns the Earth's orbital path. Much exaggerated in the diagram, this varies from almost circular to slightly more elliptical. In other words, the path stretches, varying the distance between the Earth and Sun at perihelion and aphelion. Starting at any date, it takes about 95000 years for the orbit to move through the full cycle and return to its initial path. Clearly, a variation in the distance between the Earth and Sun affects the intensity of radiation received at the Earth's surface and, therefore, the climates of Earth.

The second cycle occurs because the axis wobbles, describing a circle, rather like a toy gyroscope (see Figure 2.27). It is this wobble, due to gravitational attraction, mainly from the Sun and Moon, that causes the position of the equinoxes to move westward, taking 25800 years to complete one orbit. The phenomenon is called the precession of the equinoxes. At present

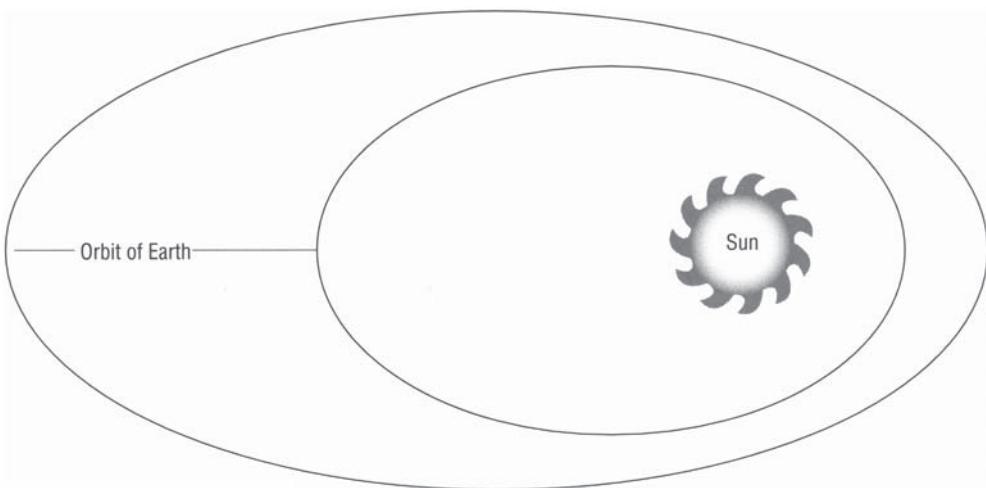


Figure 2.26 Orbital stretch
Period: about 95000 years

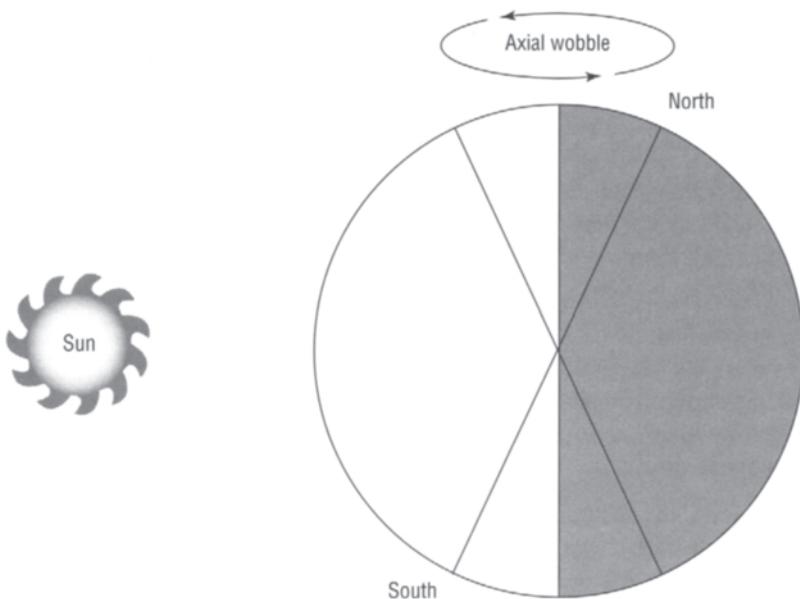


Figure 2.27 Wobble of the Earth's axis

we are at perihelion in January. In AD 15000, one half-cycle from now, we will reach perihelion in June.

The third cycle, called the ‘obliquity of the ecliptic’, relates to the angle of the Earth’s rotational axis to the ecliptic, the plane of the Earth’s orbit. Imagine the rotational axis as a straight rod projecting at both ends and forming an angle with the ecliptic. At present, that angle is 66.5° and, therefore, the axis is 23.5° from the vertical (90° to the ecliptic). Over a cycle of about 41000 years this angle varies by about 1.5° about a mean of 23.1° (see Figure 2.28).

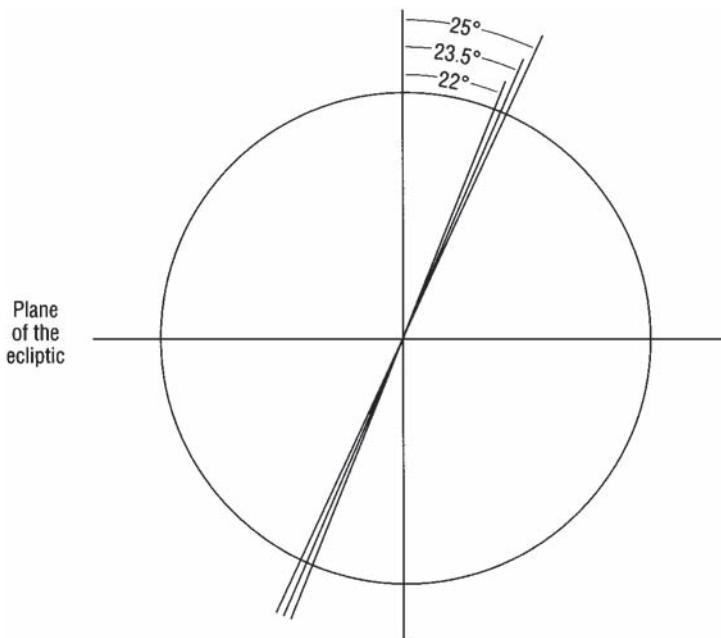


Figure 2.28 Variations in axial tilt (obliquity of the ecliptic) in degrees

These axial variations alter the area illuminated by the Sun. If the axis were at right angles to the ecliptic, for example, giving an obliquity of 0° , the half of the Earth facing the Sun would be lit evenly. Day and night would always be the same length and there would be no seasons. Tilt the Earth much more, on the other hand, say to an obliquity of 60° , and over almost the whole of each hemisphere the Sun would never set in summer or rise in winter.

Dramatic climate change occurs when the three cycles coincide, and the Milankovich solar radiation curve, which combines the three, is used to make long-term climatic predictions (deschutes.gso.uri.edu/~rutherford/milankovitch.html). It is this that allows climatologists to assert that a cooling trend which began about 6000 years ago will continue, leading us into a new ice age (HIDORE AND OLIVER, 1993, pp. 370–371), although the solar influence may be overridden by that of greenhouse gases if these continue to accumulate in the atmosphere.

In the shorter term, the solar output itself also varies. The first person to relate this to climate change was the British astronomer Edward Walter Maunder (1851–1928). Like many astronomers, he was interested in sunspots (es.rice.edu/ES/humsoc/Galileo/Things/sunspots.html), dark ‘blotches’ on the surface of the Sun that come and go in a cycle of about 11 years. Checking through old records of sunspot activity, in 1893 he discovered that very few sunspots were reported during a period of 70 years from 1645 to 1715, and for 32 years, it seems, there were no sunspots at all. He published a paper describing his findings in 1894, but it attracted little attention, any more than did earlier papers challenging the idea of the constancy of solar output, published by Maunder and by the German astronomer Gustav Spörer (1822–95). Today, the period during which sunspots were much reduced in number is known as the ‘Maunder minimum’. Its significance extends far beyond the realms of solar astronomy, because the 1645–1715 minimum Maunder identified coincides with the peak of the ‘Little Ice Age’, when average temperatures were about 1°C lower than they had been previously (LAMB, 1995, pp. 69 and 321). More recently, the American solar astronomer John A. Eddy checked the Maunder and Spörer findings, added more of his own, and found a correspondence between solar

activity and climate so close he described it ‘almost that of a key in a lock’, extending to 3000 BC (EDDY, 1977).

Again, the solar influence may be overwhelmed by that from greenhouse gases. David Thomson, a skilled statistician, has analysed data since 1659 and concluded that global temperatures are now linked more closely to atmospheric carbon dioxide concentrations than to sunspot activity or orbital effects (THOMSON, 1985), although his interpretation has been questioned by some climatologists, who think it too simple (KERR, 1995). The idea is now gaining ground that present changes in the atmosphere and climate are more likely to be due to changes in solar output and volcanic eruptions than to human intervention (CALDER, 1999).

Debate will continue for some time over what is forcing present climate change, but at least in the past it has clearly been triggered by astronomical events, and when the climate changes it can do so very quickly. At one time it was thought that ice ages begin and end gradually, it taking centuries or longer for the ice sheets to spread. This may be incorrect. According to the ‘snowblitz’ theory, a slight fall in summer temperatures in high latitudes might allow some of the winter snow to survive where in previous years it had melted. The affected areas would then be white, when previously they had been dark, thus increasing albedo and lowering temperature further. In succeeding years, the snow-covered area would increase and temperatures would continue to fall, climatic forcing by the increased albedo accelerating the change by a strongly positive feedback. It might take very little time to move from our present interglacial climates to a full glaciation. Warming can also proceed rapidly, the change from glacial to interglacial perhaps taking no more than a few decades.

Stability of the polar ice sheets

If the polar ice caps were to melt, the volume of water released into the oceans would be sufficient to raise sea levels substantially. The stability of the ice caps is therefore of great importance and their condition is monitored closely.

The ice caps comprise three major ice sheets: in Greenland, West Antarctica, and East Antarctica. The Greenland ice cap is growing thicker in some areas, thinner in others, and is shrinking slightly overall. The reduction in its size is due to the rate of flow of its outflow glaciers and is not thought to be due to climatic change.

In Antarctica, the ice sheet on the eastern side of the Transantarctic Mountains is about twice the size of that on the western side. The East Antarctic ice sheet is very firmly grounded on the underlying rock. Its size remains constant and there is not considered to be any risk of it decreasing in thickness.

The West Antarctic ice sheet is less firmly grounded and the line marking the edge of the grounded sheet is retreating. It is doing so very slowly, at a constant rate, and has been retreating at this rate for about 7500 years. The retreat is due to the way glaciers within the ice sheet are moving and not to climatic change.

Evidence from the past indicates that despite minor fluctuations, the climate throughout the present interglacial, the Flandrian, has been very stable. During the last two glaciations and the Eemian Interglacial separating them, temperatures rose and fell rapidly, by 3°C or more, bringing warmer or

cooler periods lasting several centuries or a few thousand years. (GROOTES *ET AL.*, 1993). These oscillations have since been linked to changes in ocean circulation (ZAHN, 1994).

Ancient climates are reconstructed mainly from evidence obtained from ice cores, those referring to the Eemian Interglacial and the glaciations to either side having been obtained from Greenland. Ice sheets form by the compaction of snow under the weight of overlying snow, so the ice forms in seasonal layers that can be dated by counting, much like tree rings. Temperatures are inferred by oxygen-isotope analysis. There are three isotopes of oxygen, ^{16}O , ^{17}O , and ^{18}O , but only ^{16}O and ^{18}O are of importance in climatic studies. Being lighter, water containing ^{16}O evaporates more readily than H^{18}O , so fresh water is enriched in ^{16}O as compared with sea water. The degree of enrichment depends on the temperature at which the water evaporated, because the higher the temperature, the greater the rate of evaporation and the more H^{18}O that enters the air with the H^{16}O . This allows mean surface temperature to be calculated from analyses of the ratio of $^{16}\text{O}:\text{H}^{18}\text{O}$ in dated samples of ice trapped in cores as ‘fossil precipitation’, the present ratio of $^{18}\text{O}:^{16}\text{O}=1:500$ providing a standard.

Astronomical climate forcing can be predicted, but volcanic eruptions are wholly unpredictable, at least at present. Some eruptions, but not all, have a climatic influence, although its scale is small and it is of short duration. If it is to affect climate, a volcanic eruption must inject material into the stratosphere, where it will remain for some time; tropospheric material is adsorbed on to surfaces or removed by precipitation in a matter of hours, days, or at most weeks. The eruption should also be in a low latitude. The convection cells governing the movement of low-latitude air allow only minor exchanges of tropospheric air between the northern and southern hemispheres. Stratospheric air is less affected and there is some interchange. Material injected into the stratosphere near the equator will be carried around the Earth and may also spill into higher latitudes in both hemispheres.

On 15 June 1991, the eruption of Mount Pinatubo on the island of Luzon, in the Philippines (latitude 15° N) caused the greatest stratospheric perturbation this century. The plume reached a height of about 30 km and released into the stratosphere some 30 million tonnes of aerosol composed of sulphuric acid and water. Within 14 days the material had spread across the equator, to about 10° S, and carried westward; within 22 days it had circled the planet. Eventually it spread as a blanket between about 30° N and 20° S. The presence of so much fine-particulate matter in the upper atmosphere increased the planetary albedo and thus reduced the amount of solar radiation reaching the surface, with the result that surface temperatures were depressed during the remainder of 1991 and for the whole of 1992; it was 1993 before they began to recover. In 1992, the mean global temperature was 0.2°C lower than the 1958–91 average and it would have been lower still were it not for the warming influence of the 1992 ENSO event. The eruption ended the run of warm years. Because the aerosol engaged in chemical reactions, the eruption also contributed to the greatest depletion of stratospheric ozone recorded up to that time (MCCORMICK *ET AL.*, 1995).

Mount Pinatubo was the biggest eruption this century, but it was not the only one. Five other eruptions were large enough to have had some climatic effect: those of Katmai (1912), Agung (1963), Fuego (1974), El Chichón (1982), and Cerro Hudson (1991), releasing 20, 16–30, 3–6, 12, and 3 million tonnes of aerosol respectively. In the last century there were two even larger eruptions, of Tambora (1815) and Krakatau (1883); these released more than 100 and about 50 million tonnes of aerosols respectively. The year 1815 was known as ‘the year with no summer’ and in Britain the summers of 1816 and 1817 were also wet and cold; the 1816 harvest was disastrous and there were food riots (STRATTON AND BROWN, 1978).

Our climate is changing constantly, driven by factors over which we have no hope of control. It is affected by cyclical variations in the Earth’s orbit and rotation and apparently erratic fluctua-

tions in solar output. Volcanic eruptions can depress surface temperatures and ENSO events enhance them. It may be that emissions of greenhouse gases are now overwhelming these natural forcing factors, but this does not remove them: predictions of future climate must take them into account, inherently unpredictable though some of them may be. Those attempting predictions must also bear in mind the possibility that once climate begins to change the rate of change may accelerate dramatically and that we seem to be living in unusually stable times. Predictions are concerned with the future, of course, but they must incorporate evidence gleaned from the past. Palaeoclimatologists, who study ancient climates, supply information that is vitally important to forecasters.

21 Climatic regions and floristic regions

Climates can be classified. At the simplest level, latitude, proximity to the ocean, and the convective cells transporting warm air away from the equator and cool air away from the poles provide a basic classification. Equatorial regions are warm and humid, subtropical regions, where dry air descends, are warm and dry, polar regions are cold and dry, and the mid-latitudes are mild and humid or dry with temperature extremes according to whether they are maritime or continental.

Unfortunately, it is not quite so simple as it sounds, because ‘warm’, ‘cool’, ‘dry’, and ‘humid’ are relative terms that mean little by themselves. Aridity, for example, depends not on annual precipitation, but on ‘effective precipitation’, which is precipitation minus evaporation, this being what determines the amount of moisture reaching the ground water. This, in turn, is related to temperature and a figure for the average annual temperature may conceal a very wide difference between summer and winter. Many attempts have been made to base a classificatory system on the general circulation of the atmosphere, the earliest dating from the 1930s.

The most successful scheme of this kind, proposed in 1950 by the German climatologist H. Flöhn, is illustrated in Figure 2.29. Flöhn took account of the global wind belts and distribution of precipitation (BARRY AND CHORLEY, 1982, pp. 358–373). In 1969, A.N. Strahler proposed an even simpler system based on the air masses which produce climates, dividing all climates into three types: low-latitude; mid-latitude; and high-latitude. These were subdivided according to variations in temperature and precipitation to produce 14 regional types, with a separate category for upland climates.

The two most widely used classifications, however, were introduced between 1900 and 1936 by the Russian-born German climatologist Wladimir Peter Köppen (1846–1940) and in 1931, with important revisions in 1948, by the American climatologist C. Warren Thornthwaite (1899–1963). The Köppen classification is widely used by geographers, that of Thornthwaite by climatologists.

Köppen took account of the distribution of vegetation, based originally on studies published in the last century by Alphonse de Candolle (1806–93), whose *Géographie botanique raisonnée* (1855) considered the geographical distribution of plants in relation to their physiology. From this it emerges that a summer temperature of 10°C marks the limit of tree growth, a winter temperature above 18°C is necessary for some tropical plants, and if the average winter temperature is below -3°C there will be at least some snow cover. Using these criteria and records of monthly average temperatures, Köppen defined six climatic types. In tropical rainy climates temperatures are above 18°C throughout the year; in warm, temperate, wet climates temperature in the coldest month is -3–18°C; in cold boreal-forest climates temperature in the coldest month is below -3°C and in the warmest month above 10°C; in tundra climates temperature in the warmest month is 0–10°C; in polar climates the temperature never rises above 0°C; and a final category

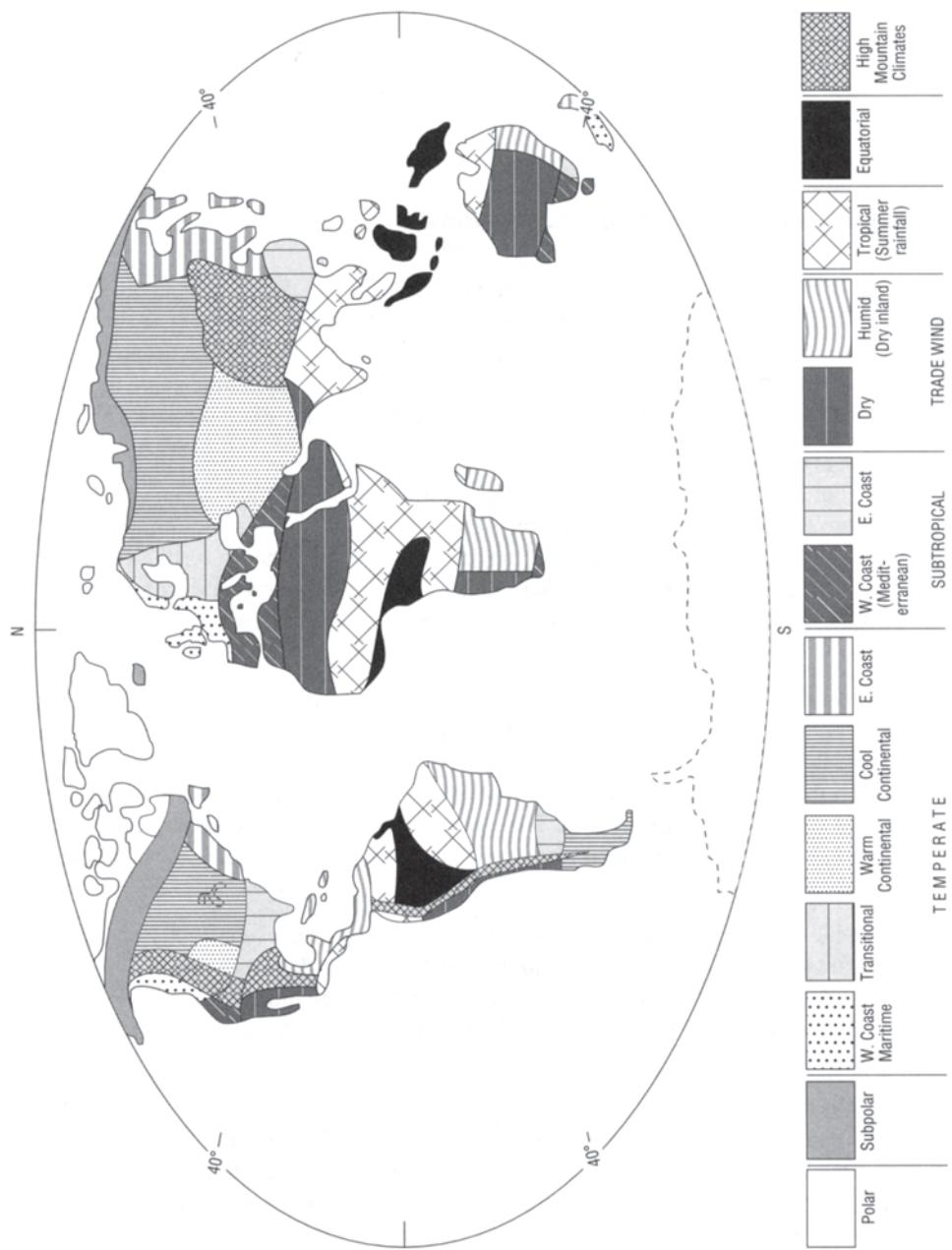


Figure 2.29 **World climate types**

of dry climates is defined by aridity. These main types were then subdivided into more detailed categories, allowing for climates with or without dry and rainy seasons, monsoon climates, and others. The relationship between temperature and plant distribution is imprecise, however, so the categories are somewhat arbitrary, with many exceptions, and his classification is rather crude, despite its popularity.

Thorntwaite adopted a different approach derived from the water required by farm crops (ALLABY, 1992a, p. 109) and based on precipitation efficiency and thermal efficiency. Both of these can be calculated. Precipitation efficiency is measured for each month as the ratio of precipitation to temperature to evaporation (as $115(r/t - 10)^{10/9}$, where r is the mean monthly rainfall in inches and t is the mean monthly temperature in °F), the sum of the 12 monthly values giving a precipitation efficiency ($P-E$) index. Thermal efficiency is calculated each month as the extent to which the mean temperature exceeds freezing (as $(t - 32)/4$); the thermal efficiency ($T-E$) index is the sum of the monthly values. The major change Thorntwaite introduced to his scheme in 1948 concerned the importance of transpiration by plants. Combined with evaporation (in practice the two cannot be measured separately in the field) this is evapotranspiration or, if water is available in unlimited amounts, ‘potential evapotranspiration’ (PE). It is calculated in centimetres from the mean monthly temperature in °C, corrected for changing day length.

Using his three indices, Thorntwaite defined nine ‘humidity provinces’ and nine ‘temperature provinces’, the respective index value doubling between each province and the next in the hierarchy. He then added further subdivisions to reflect the distribution of precipitation through the year, leading to 32 distinct climate types. The humidity provinces, with their denoting letters, are: perhumid (A); humid (B_4 , B_3 , B_2 , B_1); moist subhumid (C_2); dry subhumid (C_1); semi-arid (D); and arid (E). The temperature provinces are: frost (E’); tundra (D’); microthermal (C'_1 , C'_2); mesothermal (B'_1 , B'_2 , B'_3 , B'_4); and megathermal (A’). This classification makes no assumptions about the distribution of plants, but is based wholly on recorded data.

These classifications are described as ‘empirical’, because they are based on data. Their disadvantage arises from the fact that divisions among sets of continuous variables are inevitably arbitrary, so the number of categories is potentially huge, and the more regional variations a scheme recognizes the more unwieldy it becomes. ‘Genetic’ classifications, derived from seasonal patterns of insolation and precipitation or the dominant air masses, are not widely used, but there are several of them. Indeed, there are many classificatory systems (HIDORE AND OLIVER, 1993, pp. 263–264), but those of Köppen and Thorntwaite remain the most popular.

Thorntwaite devised a scheme to classify climates independently of the vegetation each type supports, but the historical association between climate classification and plant distribution is close. Up to a point the link is obvious. Tropical rain forests flourish in the humid tropics, cacti and succulents in arid climates, conifer forests in high latitudes, and tundra vegetation borders the barren polar regions. Clearly, plants occur only where the climate suits them; bananas do not grow in Greenland (at least, not in the open). Although plant distribution is linked to climate, however, other factors also influence it. Continental drift has separated what were once adjacent landmasses supporting similar plants, producing very discontinuous distributions. The southern beeches (*Nothofagus*), for example, occur in Australasia and western South America, and pepper bushes (*Clethra*) in China and South-East Asia and from the south-eastern United States to the northern and central regions of South America, but with fossil remains in Europe. Major climate changes alter vegetation patterns, but often leave remnants of the former pattern surviving as isolated relicts. The strawberry tree (*Arbutus unedo*) belongs to a pattern of plants known as Lusitanian; these occur in south-western Europe, but also, as relicts, in southern Ireland and Brittany.



Figure 2.30 Floristic regions

- | | |
|---------------------------|---------------------------|
| 1 Arctic | 11 Sudanian-Sindian |
| 2 Euro-Siberian | 12 Ethiopian |
| 3 Irano-Turanian | 13 West African |
| 4 Sino-Japanese | 14 East African |
| 5 Mediterranean | 15 South African |
| 6 Hudsonian | 16 Madagascan |
| 7 Pacific North American | 17 Indian |
| 8 Atlantic North American | 18 South-East Asian |
| 9 Macaronesian | 19 Malaysian-Papuan |
| 10 Saharo-Sindian | 20 New Caledonian |
| | 21 Fijian Pacific |
| | 22 Polynesian Pacific |
| | 23 Hawaiian |
| | 24 Central American |
| | 25 Pacific South American |
| | 26 Parano-Amazonian |
| | 27 Argentinian |
| | 28 Australian |
| | 29 New Zealand |
| | 30 South Oceanic |

Nevertheless, regions of the world can usually be defined in terms of the plants occurring naturally within them and those regions coincide, more or less, with the climatic zones. The plants growing in a particular area comprise the ‘flora’ of that area and floras can be grouped into units, called ‘phytochoria’ (singular *phytochorion*), in which small unrelated floras, such as the Lusitanian in northern Europe, are designated ‘elements’. Once defined, phytocoria can then be grouped further into a hierarchical system. The highest category is the floral realm or kingdom (both terms are used), which is divided into regions. Regions are subdivided further into provinces or domains (the terms are synonymous), each comprising a number of districts. Some classifications allow intermediate ranks and subdivisions of districts (MOORE, 1982, pp. 210–219). Realms are identified by the presence of particular plant families, regions by the presence of 20–30 per cent of plant genera that are not found elsewhere (i.e. endemic genera), and provinces by their endemic genera.

Most classifications recognize four floral realms: Holarctic, Palaeotropic, Neotropical, and Austral. The Holarctic Realm comprises North America, Greenland, Europe, and Asia except for India and the south-west and south-east (which became attached to the main landmass during the Tertiary). Floristically, the mountains extending from the Atlas range in North Africa across southern Asia to the Himalayas mark the southern boundary of the northern hemisphere in the Old World. With a few exceptions, coniferous trees occur north of the boundary and palms to its south.

The Palaeotropic Realm (the name means ‘old’ tropical) comprises Africa south of the Atlas Mountains except for southern Africa, Madagascar, Arabia, southern Asia including India, and the islands of the tropical Pacific. The Neotropical Realm (‘new’ tropical) comprises Central America including the southern tips of California and Florida, the Caribbean, and most of South America. Although their climates are similar, floristically these tropical realms differ from one another markedly because of the time that has elapsed since continental drift separated them. Cacti, for example, are characteristic of the New World and are one of the defining families (Cactaceae) of the Neotropical Realm; those found in the Old World have been introduced. This is why they are regarded as two distinct realms, rather than one.

The Austral (or Southern) Realm comprises the southern part of South America, southern Africa, Australia, New Zealand, and the islands of the southern Atlantic and Pacific. Here, too, the landmasses are now isolated from one another. Southern Africa and southern South America differ floristically from the rest of the continents to their north and share some plant families with Australia and New Zealand. On this basis they are grouped together as one floral realm or, in some classifications, ranked as individual Australian, Cape, and Antarctic Realms. Figure 2.30 shows these realms and the 30 floral regions of which they are composed.

Floristic realms and regions vary in size, but all are vast and difficult to comprehend. It is not until their subdivision reaches the provincial level that they become easily recognizable. Western Europe, for example, from northern Spain to Denmark and the Norwegian coast, constitutes the Atlantic Province. The Boreal Province, supporting vast tracts of coniferous forest (known in Russia as the taiga) forms a belt across Europe and Asia between the Ural River and Gulf of Finland and latitude 60° N. The North American equivalent, covering most of Alaska and Canada south of the Arctic, is called the Hudsonian Province.

Animal distribution is also described geographically and, because particular animals are often associated with particular plants, zoographical and floristic realms almost coincide. The concept of realms, with their subdivisions, should not be confused with that of biomes, which are defined ecologically. Floristic classification reflects climates, past and present, and the history as well as present geography of the planet.

End of chapter summary

All living organisms, including us, depend entirely on the materials from which the physical Earth is made and on the energy our planet receives from the Sun. It is important, therefore, to understand, at least in a very general way, how rocks and minerals form and how the landscape changes over long periods of time.

Ocean currents and all atmospheric movements are driven by energy the Earth receives from the Sun in the form of electromagnetic radiation. The movement of air and water produce the climates of the world, and the weather we experience day by day. Climate is the principal factor in determining the plants and animals that live in a particular area.

Climate is not constant, however. It has changed many times in the past. At present we are living in an interglacial and one day the ice sheets and glaciers may begin the advance that marks the dawn of a new ice age. Meanwhile, many people fear we may be inadvertently modifying the climate ourselves.

End of chapter points for discussion

Why are the oceans so important climatically?

What evidence is there that glaciations occurred?

Why are there deserts in the tropics of both hemispheres?

How does climate affect the distribution of plants?

See also

Formation and structure of the Earth (section 6)

Formation of rocks, minerals, and geologic structures (section 7)

Weathering (section 8)

Evolution of landforms (section 9)

Solar energy (section 11)

Albedo and heat capacity (section 12)

Greenhouse effect (section 13)

Structure of the atmosphere (section 14)

Atmospheric circulation (section 15)

Ocean circulation (section 16)

Weather and climate (section 17)

Glacials, interglacials, interstadials (section 18)

Dating methods (section 19)

Climate change (section 20)

Fresh water (section 22)

Soils (section 26)

Transport by water and wind (section 27)

Soil, climate, and land use (section 28)

Mining for fuels (section 30)

Mining and processing of minerals (section 31)

Biogeography (section 32)

Nutrient cycles (section 33)

Transnational pollution (section 62)

Further reading

Air: The Nature of Atmosphere and the Climate. Michael Allaby. 1992. Facts on File, New York. A popular science book providing useful information on atmosphere and climate.

Climatology: An Atmospheric Science. John J. Hidore and John E. Oliver. 1993. Macmillan, New York. Provides a broad overview of its subject, clearly written in simple language.

The Holocene: An Environmental History. Neil Roberts. 1989. Basil Blackwell, Oxford. Describes briefly the history of the last 10000 years, with details of the methods used to interpret the past.

Principles of Physical Geology. Arthur Holmes. 2nd edn 1965. Nelson, Walton-on-Thames. Old, but still possibly the most approachable introduction to the earth sciences.

The Study of Landforms. R.J.Small. 1970. Cambridge University Press, Cambridge. Provides a broad general introduction to geomorphology.

Water. Michael Allaby. 1992. Facts on File, New York. A popular science book with useful explanations of climate and weather.

Notes

- 1 These are outlined in Hutchison, R. 1981. 'The origin of the Earth', in Cocks, L.R.M. (ed.) *Evolving Earth: Chance, change and challenge*. British Museum (Natural History) and Cambridge Univ. Press, London, pp. 5–16.
- 2 There is a brief outline of the current ideas of the mechanism underlying movements in the mantle in 'Making waves', by Sarah Simpson, *Scientific American*, August 1999, p. 10.
- 3 There is a brief description of the wildlife on Surtsey in Helfferich, Carla. 1993. 'Sandwort, Seabirds, and Surtsey', Article 1132, Alaska Science Forum. www.gi.alaska.edu/ScienceForum/ASF11/1132.html.
- 4 For a rather technical explanation see Thorpe, Richard and Brown, Geoff. 1985. *The Field Description of Igneous Rocks*. Geological Society of London Handbook. Open Univ. Press, Milton Keynes (published by John Wiley and Sons in the US and Canada).
- 5 The formation of Britain is described fairly simply in Dunning, F.W., Mercer, I.W., Owen, M.P., Roberts, R.H., and Lambert, J.L.M. 1978. *Britain Before Man*. HMSO for the Institute of Geological Sciences, London.
- 6 For a guide to the interpretation of sedimentary rocks, see Tucker, Maurice E. 1982. *The Field Description of Sedimentary Rocks*. Geological Society of London Handbook. Open Univ. Press, Milton Keynes (published by John Wiley and Sons in the US and Canada).
- 7 See Fry, Norman. 1984. *The Field Description of Metamorphic Rocks*. Geological Society of London Handbook. Open Univ. Press, Milton Keynes (published by John Wiley and Sons in the US and Canada).
- 8 These are described in Angel, Heather, Duffey, Eric, Miles, John, Ogilvie, M.A., Simms, Eric, and Teagle, W.G. 1981. *The Natural History of Britain and Ireland*. Michael Joseph, London, pp. 108–110.
- 9 Stefan's law is expressed as: $E=\sigma T^4$, where E is the amount of radiation emitted, T is the temperature, and σ is the Stefan-Boltzmann constant, which is the amount of radiant energy released by a black body. It is represented by σ and is equal to $5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ (watts per square meter per kelvin to the fourth power). The temperature is in kelvins and the energy units are watts per square meter (W m^{-2}). The law (and constant) were discovered in 1879 by the Austrian physicist Josef Stefan (1835–93), and at first they were known as Stefan's law and constant. In 1884, the Austrian physicist Ludwig Eduard Boltzmann (1844–1906), Stefan's former student, showed that the law holds only for black bodies and his name was added to that of the law and constant.
- 10 Wien's law can be stated as $\lambda_{\max} = C/T$, where λ_{\max} is the wavelength of maximum emission, C is Wien's constant, and T is the temperature in kelvins. Wien's constant is $2897 \times 10^{-6} \text{ m}$ ($2897 \mu\text{m}$), so the law becomes: $\lambda_{\max} = (2897/T) \times 10^{-6} \text{ m}$. It is valid only for radiation at short wavelengths. The law was discovered in 1896 by the German physicist Wilhelm Wien (1864–1928) and for it he was awarded the 1911 Nobel Prize for physics.

References

- Adams, Peter. 1977. *Moon, Mars and Meteorites*. HMSO for the British Geological Survey.
- Allaby, Ailsa and Michael (eds) 1999. *The Concise Oxford Dictionary of Earth Sciences*. 2nd edition. Oxford University Press, Oxford.
- Allaby, Michael. 1992. *Elements: Air*. Facts on File, New York.
- . 1992a. *Elements: Water*. Facts on File, New York.

- . 1998. *Dangerous Weather*. Floods. Facts on File, Inc., New York.
- Avery, Dennis. 1995. ‘Saving the Planet with Pesticides: Increasing Food Supplies While Preserving the Earth’s Biodiversity’, in Bailey, Ronald (ed.) *The True State of the Planet*, The Free Press, New York, pp. 73–76.
- Balling, Robert C., Jr. 1995. ‘Global Warming: Messy Models, Decent Data and Pointless Policy’, in Bailey, Ronald (ed.) *The True State of the Planet*, The Free Press, New York, pp. 83–107.
- Barry, Roger G. and Chorley, Richard J. 1982. *Atmosphere, Weather and Climate*. Methuen, London.
- Boehmer-Christiansen, Sonja A. 1994. ‘A scientific agenda for climate policy?’, *Nature*, 372, 400–402.
- Bolin, Bert. 1995. ‘Politics of climate change’ (letter in reply to Boehmer-Christiansen), *Nature*, 374, 208.
- Bowler, Peter J. 1992. *The Fontana History of The Environmental Sciences*. Fontana Press, London.
- Burbank, Douglas W. 1992. ‘Causes of recent Himalayan uplift deduced from deposited patterns in the Ganges basin’, *Nature*, 357, 680–682.
- Calder, Nigel. 1999. ‘The Carbon Dioxide Thermometer and the Cause of Global Warming’, *Energy and Environment*, 10, 1, 1–18.
- Cane, Mark A., Eshel, Gidon, and Buckland, R.W. 1994. ‘Forecasting Zimbabwean maize yield using eastern equatorial Pacific sea surface temperature’. *Nature*, 370, 204–205.
- Clark, John. 1977. *Coastal Ecosystem Management*. Wiley-Interscience, New York.
- Cowling, Sharon A. 1999. ‘Plants and Temperature-CO₂ Uncoupling’, *Science*, 285, 1500–1501.
- Dalziel, Ian W.D. 1995. ‘Earth before Pangaea’, in *Scientific American*, January, pp. 38–43.
- Eddy, John A. 1977. ‘The case of the missing sunspots,’ *Scientific American*. pp 80–89.
- Gentry, A.W. and Sutcliffe, A.J. 1981. ‘Pleistocene geography and mammal faunas’, in Cocks, L.R.M. (ed.) *The Evolving Earth: Chance, change and challenge*. British Museum (Natural History) and Cambridge Univ. Press, London, pp. 237–251.
- Gleick, James. 1988. *Chaos: Making a new science*. Heinemann, London, pp. 9–31.
- Graham, A.L. 1981. ‘Plate tectonics’, in Cocks, L.R.M. (ed.) *The Evolving Earth: Chance, change and challenge*. British Museum (Natural History) and Cambridge University Press, London. Chapter 11.
- Grootes, P.M., Stuiver, M., White, J.W.C., Johnsen, S., and Jouzel, J. 1993. ‘Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores,’ *Nature*, 366, 552–554.
- Hambrey, M.J. and Harland, W.B. 1981. ‘The evolution of climates’, in Cocks, L.R.M. (ed.) *The Evolving Earth: Chance, change and challenge*. British Museum (Natural History) and Cambridge University Press, London. Chapter 9.
- Harvey, John. G. 1976. *Atmosphere and Ocean*: Our fluid environments. Artemis Press, Sussex.
- Hidore, John J. and Oliver, John E. 1993. *Climatology: An atmospheric science*. Macmillan, New York.
- Holm, Søren and Harris, John. 1999. ‘Precautionary principle stifles in discovery’, letter to *Nature*, 400, 398.
- Holmes, Arthur. 1965. *Principles of Physical Geology*. Nelson. Walton-on-Thames. 2nd edition.
- Hudson, Norman. 1971. *Soil Conservation*. B.T.Batsford Ltd, London.
- Hunt, Charles B. 1972. *Geology of Soils: Their evolution, classification, and uses*. W.H.Freeman and Co., San Francisco.
- Intergovernmental Panel on Climate Change. 1992. *1992 IPCC Supplement: Scientific assessment of climate change*. WMO and UNEP.
- Jacobs, G.A., Hurlbut, H.E., Kindle, J.C., Metzger, E.J., Mitchell, J.L., Teague, W.J., and Wallcraft, A.J. 1994. ‘Decade-scale trans-Pacific propagation and warming effects of an El Niño anomaly’, *Nature*, 370, 360–363.
- Johnson, Nicholas and David, Andrew. 1982. ‘A Mesolithic site on Trevose Head and contemporary geography’, in *Cornish Archaeology*, 21, 67–103.
- Kempe, D.R.C. 1981. ‘Deep ocean sediments’, in Cocks, L.R.M. (ed.) *The Evolving Earth: Chance, change and challenge*. British Museum (Natural History) and Cambridge University Press, London, pp. 116–117.

- Kendeigh, S.Charles. 1974. *Ecology With Special Reference to Animals and Man*. Prentice-Hall, Eaglewood Cliffs, New Jersey.
- Kerr, Richard A. 1995. 'Sun's role in warming is discounted,' *Science*, 268, 28–29.
- Lamb, H.H. 1995. *Climate, History and the Modern World*. 2nd ed. Routledge, London.
- McCormick, M.Patrick, Thomason, Larry W., and Trepte, Charles R. 1995. 'Atmospheric effects on the Mt Pinatubo eruption,' *Nature*, 373, 399–404.
- Moore, David M. (ed.) 1982. *Green Planet: The story of plant life on Earth*. Cambridge University Press, Cambridge.
- Morell, Virginia. 1995. 'The earliest art becomes older—and more common', *Science*, 267, 1908–1909.
- Pennington, Winifred. 1974. *The History of British Vegetation*, 2nd ed. Hodder and Stoughton, London.
- Raffensperger, Carolyn, Tickner, Joel, Schettler, Ted, and Jordan, Andrew. 1999. '...and can mean saying "yes" to innovation', letter to *Nature*, 401, 207.
- Raven, Peter H., Berg, Linda R., and Johnson, George B. 1993. *Environment*. Saunders College Publishing, Orlando, Florida.
- Roberts, Neil. 1989. *The Holocene: An environmental history*. Basil Blackwell, Oxford.
- Simmons, I.G., Dimbleby, G.W., and Grigson, Caroline. 'The Mesolithic', in Simmons, Ian and Tooley, Michael (eds) 1981. *The Environment in British Prehistory*. Duckworth, London.
- Small, R.J. 1970. *The Study of Landforms*. Cambridge Univ. Press.
- Stratton, J.M. and Brown, Jack Houghton. 1978. *Agricultural Records AS 22–1977*, 2nd edn. John Baker, London.
- Thomson, David J. 1985. 'The seasons, global temperature, and precession,' *Science*, 268, 59–68.
- Tolba, Mostafa K. and El-Khody, Osama A. 1992. *The World Environment 1972–1992*. UNEP and Chapman and Hall, London.
- Trimble, Stanley W. 1999. 'Decreased Rates of Alluvial Sediment Storage in the Coon Creek Basin, Wisconsin, 1975–93', *Science*, 285, 1244–1246.
- Wagner, Frederike, Bohncke, Sjoerd J.P., Dilcher, David L., Kürschner, Wolfram M., van Geel, Bas, and Visscher, Henk. 1999. 'Century-Scale Shifts in Early Holocene Atmospheric CO₂ Concentration', *Science*, 284, 1971–1973.
- Windley, Brian F. 1984. *The Evolving Continents*. John Wiley and Sons, Chichester.
- Zahn, Rainer. 1994. 'Linking ice-core records to ocean circulation,' *Nature*, 371, 289.

3

Physical Resources

When you have read this chapter you will have been introduced to:

- the hydrologic cycle
- the life cycle of lakes
- salt water, brackish water, and desalination
- irrigation, waterlogging, and salinization
- soil formation, soil ageing, and soil taxonomy
- soil transport
- soil, climate, and land use
- soil erosion and its control
- mining and processing fuels
- mining and processing minerals

22 Fresh water and the hydrologic cycle

In the sense used here, a ‘resource’ is a substance a living organism needs for its survival. There are also non-material resources, such as social contact and status, which may be essential to a feeling of well-being or even to survival itself, but these are not considered here.

Non-humans as well as humans make use of the resources available to them; animals need such things as food, water, shelter, and nesting sites, all of which are resources, as are the sunlight and mineral nutrients required by plants. Human biological requirements are similar to those of other animals. Like them, we need food, water, and shelter, although we differ from other species in the means we have developed for obtaining them. It is because human and non-human requirements often coincide that sometimes we find ourselves in direct competition for resources with non-humans. It is not only we who find crop plants edible and nutritious, for example, and before we can build houses to shelter ourselves we must clear the land of its previous, non-human occupants.

Water is, perhaps, the most fundamental of the resources we require. Without water, as the cliché has it, life could not exist on land. Our bodies are largely water (by weight), and if you add together the ingredients listed on many food packets you will find they seldom amount to more than about half the total weight: the remainder is water.

It is not any kind of water we need, of course, but fresh water. Sea water is of only limited use to us, and out of reach for people living deep inside continents, and drinking it is harmful, although it can be rendered potable by the removal of its dissolved salts. For the most part, therefore, we humans must obtain all the water we need from rivers, lakes, and underground aquifers. In the world as a whole, it is estimated that by the year 2000 we will be using about 4350 km^3 (4.35×10^{15} litres) of water a year. Of this, almost 60 per cent will be needed for crop irrigation, 30 per cent for industrial processes and cooling, and 10.5 per cent for domestic cooking, washing, and drinking (RAVEN *ET AL.*, 1993, p. 273).

Of all the water in the world, 97 per cent is in the oceans, so our freshwater needs must be met from the remaining 3 per cent. It is not even that simple, however, because of all the fresh water, more than

half is frozen in the polar icecaps and glaciers and about 0.5 per cent is so far below ground as to be beyond our reach. Atmospheric water vapour, falling rain and snow, and flowing rivers contain no more than about 0.005 per cent of the planet's water (KUPCHELLA AND HYLAND, 1986, pp. 222–223). Stated like this, the amount available to us sounds alarmingly small, but it is so only as a proportion of the total. The quantity available to us, including that in lakes and inland seas, is in the region of 15×10^{18} litres.

Water can exist as either gas or liquid at temperatures commonly encountered near the surface and consequently it is constantly evaporating and condensing again. Each year, some 336×10^{15} litres evaporates from the oceans and 64×10^{15} litres from the land surface (including water transpired by plants). About 300×10^{15} litres falls as precipitation over the oceans and 100×10^{15} litres over land, and 36×10^{15} litres flows from the land back to the sea (HARVEY, 1976, p. 22). This movement of water between oceans, air, and land constitutes the hydrologic cycle, and by dividing the quantity of water at each stage of the cycle by the amount entering or leaving, it is possible to discover approximately the average time a water molecule remains in each: its residence time. This reveals that a molecule spends about 4000 years in the ocean, 400 years on or close to the land surface, and 10 days as vapour in the atmosphere.

Most of the water falling on land evaporates again almost immediately or is taken up by plant roots and returned to the atmosphere by transpiration. Some flows directly over the surface, down slopes and into lower ground where it may enter lakes, rivers, or marshes. What remains drains downward through the soil until it encounters a layer of impermeable clay or rock, then flows laterally, very slowly, through the soil. Were it not to flow, but simply accumulate, the ground would soon be waterlogged and water would lie at the surface. Above the impermeable material a layer of soil is saturated with water. This is ground water and its upper limit, above which the soil is not saturated, is the water table. Permeable material through which ground water flows is called an 'aquifer' and it may lie deep beneath the surface. Aquifers are permeable because the particles of which they are composed, such as gravel or sand, are not packed together so tightly as to leave no spaces between them. They are said to be 'unconsolidated' and allow water to flow through them. Other aquifers are made from material, such as chalk or sandstone, which are consolidated (solid) but nevertheless have fissures, or pore spaces within their granular structure, through which water can flow.

It is obviously most convenient to obtain our supplies of fresh water from the nearest river or lake, but this may be too distant or insufficient. In that case it may be possible to obtain water from an aquifer, by sinking a borehole into it and pumping out the water. Figure 3.1 illustrates this and also shows what happens: abstraction lowers the water table around the borehole,

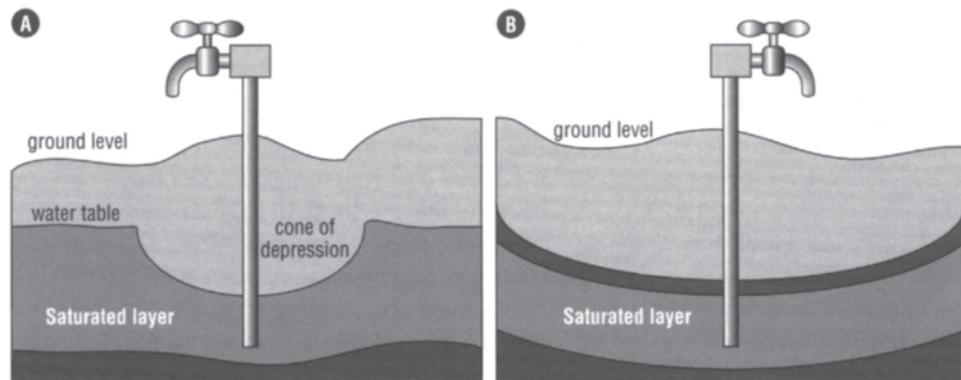


Figure 3.1 **Water abstraction**

forming a ‘cone of depression’. If the rate of abstraction exceeds that at which the aquifer is recharged, the water table will fall over a wide area, eventually to a level at which the yield from the borehole decreases and the aquifer is exhausted. In the United States, there are parts of the Great Plains, California, and southern Arizona where the severe depletion of aquifers for irrigation now threatens future water supplies and also reduces water quality. Quality is affected because, in coastal regions, as the water table falls salt water enters to recharge it, and anywhere that toxic mineral salts dissolve in ground water, reducing the volume of water may increase their concentration, so the water requires more extensive, and therefore costly, processing to render it drinkable (RAVEN *ET AL.*, 1993, pp. 279–281). Pollution of this kind is natural, although caused by human over-exploitation of a resource, but ground water can be polluted by industrial or domestic wastes.

Lowering the water table can also cause ground subsidence due to the reduction in volume of the material comprising the saturated layer as this dries. Between 1865 and 1931, groundwater abstraction in London caused the ground to subside at $0.91\text{--}1.21\text{ mm yr}^{-1}$, producing a total subsidence of $0.06\text{--}0.08\text{ m}$. In Tokyo, the ground subsided 4 m between 1892 and 1972, at a rate of 500 mm yr^{-1} , and Mexico City is sinking at $250\text{--}300\text{ mm yr}^{-1}$ for the same reason (GOUDIE, 1986, p. 207).

Not all aquifers require pumping. An unconfined aquifer is one into which water drains freely from above, but where two approximately parallel impermeable layers are separated by a layer of porous material, the resulting aquifer is said to be confined. Natural undulations in a confined aquifer produce low-lying areas in which water is under pressure from the water at a higher level to either side (see Figure 3.1B). This water will flow without pumping from a borehole drilled into the aquifer through the upper impermeable layer and it will continue to flow provided the aquifer is constantly recharged by water draining into the hollow. The result is an ‘overflowing’ or ‘artesian’ well.

Where the water table reaches the surface, water will flow spontaneously, as a spring, and on sloping ground it will form a stream and eventually, through the merging of many small streams, a mighty river. Rivers also supply water, but since long before our ancestors invented wheeled vehicles and built roads for them they have also been used to convey people and goods. It is no coincidence that most of the world’s major inland cities are located beside large rivers. Almost any river might serve as an example, but the Rhine is an especially good one, because it flows across a densely populated continent for a distance of 1320 km. Figure 3.2 shows the river together with some of its more important tributaries and the principal cities that border it.

Over the centuries the cities along the Rhine prospered and grew, and as Europe industrialized several of them became important manufacturing centres. Most industries use water and produce liquid wastes, and humans produce sewage, a mixture of urine, faeces, and water that has been used for washing and cooking. At one time all this was poured into the river, which removed it, and wastes discharged into the Rhine were joined by those discharged into its tributaries, including the Emscher, which drains the Ruhr and enters the Rhine north of Düsseldorf.

Rivers have a remarkable capacity for cleaning themselves, because their waters are continually replenished and contaminants removed by extreme dilution, precipitation and burial beneath other sediment, or, most of all, by bacterial activity that breaks down large, organic molecules into simpler, biologically harmless compounds. In the case of rivers such as the Rhine, however, transporting foul water merely delivers it to the next city downstream, where it must be treated before it can be used, and the further downstream people live, the more their drinking water will cost them. In modern times the problem has been addressed, but it was not simple. As Figure 3.3 shows, water drains into the Rhine from an area of $220\,150\text{ km}^2$ in six countries. Why should the Swiss pay more to treat effluent prior to discharge for the benefit of the distant Netherlands? Why should the French regulate discharges from their chemical industries in Alsace when the

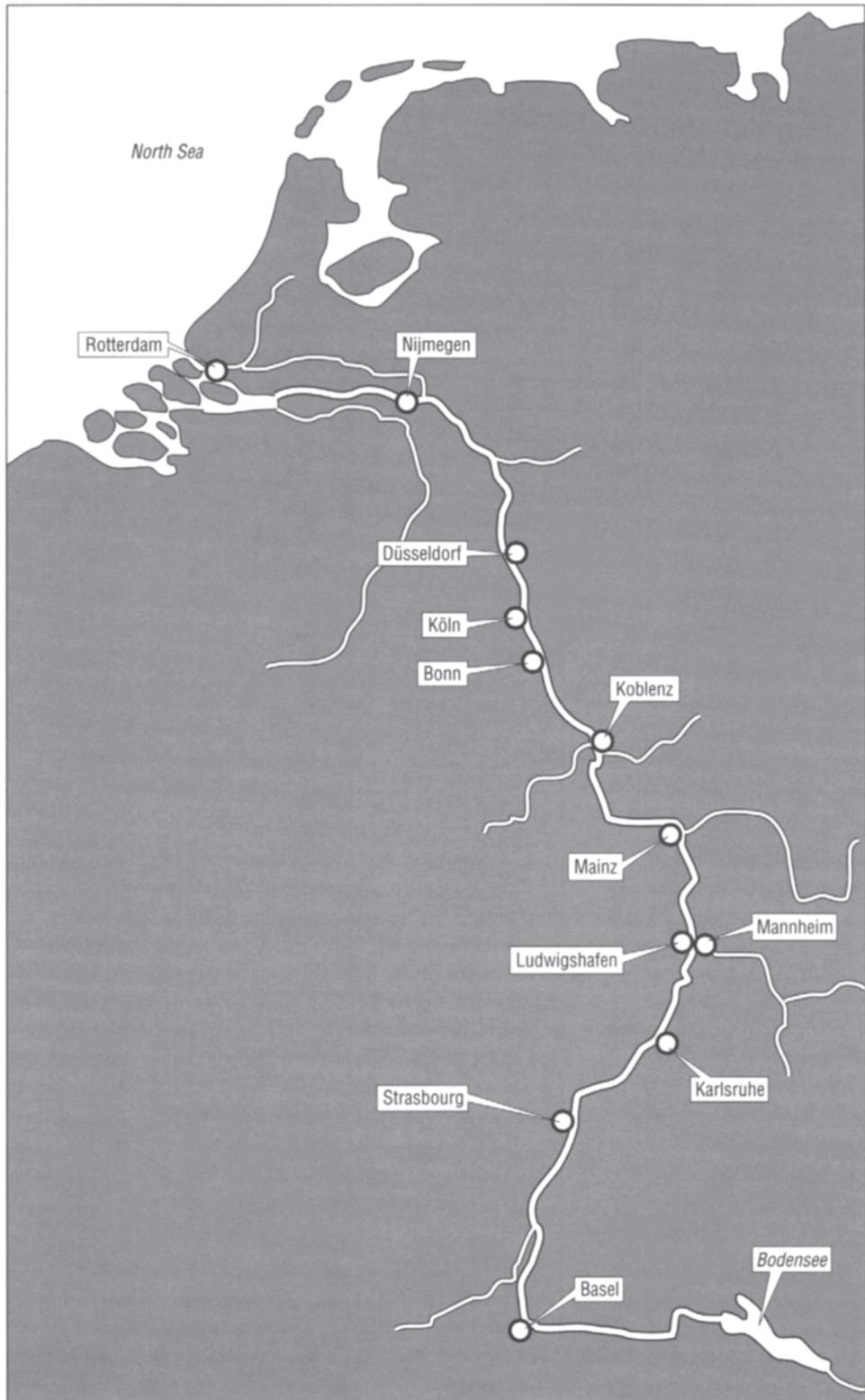


Figure 3.2 Principal cities bordering the Rhine (not to scale). Total length of the Rhine 1320 km

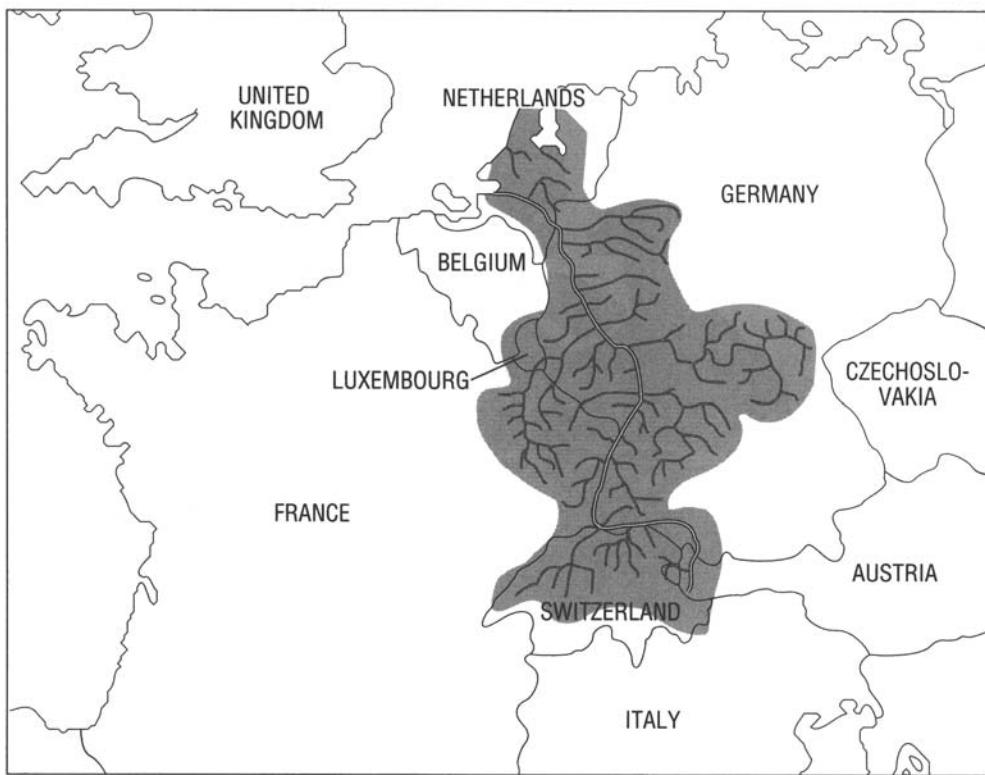


Figure 3.3 The Rhine basin, draining land in six countries

principal source of pollution was the German Ruhr? Fortunately, such transnational issues can now be resolved within the European Union, where mechanisms exist to ensure that the costs of antipollution measures are shared equitably.

Regulations are necessary, but accidents cannot be prevented by legislation and they can cause serious harm. On 1 November 1986, there was a fire at a warehouse near Basel owned by the chemical company Sandoz. Water used to fight the fire washed an estimated 30 tonnes of chemicals into the Rhine, including mercury and organophosphorus compounds and a red dye, rhodamine, that allowed the progress of the pollutants to be observed. The accident was exacerbated by a smaller spillage of herbicide on the preceding day from a Ciba-Geigy plant, also at Basel. By 12 November pollution was severe between Basel and Mainz, the river being declared ‘biologically dead’ for 300 km downstream from Basel, and by the time the affected water reached the Netherlands its mercury content, of $0.22 \mu\text{g litre}^{-1}$, was three times the usual level. Drinking water had to be brought by road to supply several cities. Despite the severity of the incident, however, the river had almost recovered one year later (MASON, 1991, pp. 2–3). Switzerland is not a member of the EU, but its government accepted responsibility for the 1986 pollution incident and promised to consider bringing its antipollution regulations into line with those of the EU (ALLABY, 1987).

Water is a so-called ‘renewable’ resource. After it has been used it returns to the hydrologic cycle and in time it will be used again. It is also abundant globally and the oceans are so vast that their capacity for absorbing, diluting, and detoxifying pollutants is immense. Despite this, the provision of wholesome fresh water and the hygienic disposal of liquid wastes in the impoverished semi-arid regions of the world is woefully inadequate. It is there that fetching water for ordinary domestic use

involves arduous hours of walking and carrying, mainly by women and children, and where debilitating water-borne diseases are common.

The resource is renewable, but distributed unevenly, and its efficient management requires an elaborate infrastructure of reservoirs, treatment plant, pipelines, and sewerage, coordinated within an overall strategy by an authority with the power to prevent abuses. For people in those regions, improvements in living standards depend crucially on the establishment of such strategies for water management, and once living standards begin to rise it is inevitable that the demand for water will increase substantially. As rising demand encounters limits in the supply available, conflicts may ensue, as they have already between Israel and Jordan over abstraction from the river Jordan. This is one of the most formidable challenges facing us. It is encouraging to note, however, that throughout history, competition between nations for scarce water resources has almost invariably been settled peacefully.

23 Eutrophication and the life cycle of lakes

In the late 1960s there was widespread popular concern over the pollution of rivers, lakes, and ground water by nitrate from sewage, farm effluents, but most of all by leaching from farmed land. It was feared that high nitrate levels in water might lead to health problems (principally methaemoglobinæmia, or ‘blue-baby’ syndrome) in infants less than 6 months old. Methaemoglobinæmia is very rare, but between 1945 and 1960 about 2000 cases were reported in the world as a whole, causing the deaths of 41 infants in the United States and 80 in Europe. The fear was not unreasonable. Today, when nitrate levels in water exceed a permitted maximum parents are advised to use bottled water for mixing infant foods and drinks. There were also fears that nitrates might form nitrous acid (HNO_3) in the body and react with amides (derived from ammonia by the substitution of an organic acid group for one (primary amide), two (secondary), or all three (tertiary) of its hydrogen atoms) or amines (also formed from ammonia, when one or more of its hydrogen atoms are replaced by a hydrocarbon group). Amines and amides are common and the product of the reaction would be *N*-nitrosamines and *N*-nitrosamides, which are known to cause cancer in experimental animals. In fact, there is no evidence that nitrate causes cancer in humans (ROYAL COMMISSION ON ENVIRONMENTAL POLLUTION, 1979, pp. 87–92). Indeed, dietary nitrates have no adverse effect whatever on human health. Although nitrites remain implicated in infant methaemoglobinæmia, it is now known that they are formed in feeding bottles by bacterial action on nitrates contained in the food in the bottle. Nitrates in the water are not involved (L'HIRONDEL, 1999). In parallel with this there was also concern that the nitrate loading of waters would cause their widespread over-enrichment (eutrophication).

Nitrogen is an essential plant nutrient and plants take it up readily in the form of nitrate (NO_3^-) ions, because all nitrates are highly soluble in water. Grass is present throughout the year, so its roots are always absorbing nitrate. Arable fields, on the other hand, are bare for part of the year, often at times of heavy rainfall. With no plant roots to intercept the nitrate, it is washed (leached) from the soil. Nitrate pollution was perceived as a problem in the 1960s because of agricultural changes that had taken place in Britain in the preceding years.

In 1938, the area of land growing arable crops in Great Britain was smaller than it had been at any time since the middle of the last century. The depression of the 1930s had so reduced the profitability of farming that large areas were almost abandoned, and as the Second World War began, with the likelihood of a sea blockade to restrict the import of food, the British people faced real hunger. Drastic steps were taken to increase agricultural output and after the war these continued as farming modernized. A major consequence of these changes was a substantial reduction in the area growing

grass and a corresponding increase in the area growing cereals. In 1938, less than 1.2 million ha was sown to barley and wheat; in 1966 those crops occupied 3.3 million ha. During the same period, the area devoted to permanent and temporary grassland fell from 8.4 million ha to 6.8 million ha. The 2.1 million ha increase in the cereal area was achieved by reducing the grassland area. (MAFF, 1968, p. 34)

Thus the change from grassland to cereal cropping led inevitably to an increase in the movement of nitrate from the soil and into surface and ground water. The widespread introduction of soluble, nitrogen-based fertilizers exacerbated the problem, especially when heavy applications were followed by very wet weather, but the fertilizer contribution should not be exaggerated. In 1964, for example, nitrogen runoff was measured following 114 mm of rain in two falls in Missouri (SMITH, 1967). Bare soil, which had received no fertilizer, lost 0.9 kg N ha⁻¹; unfertilized maize and oats lost 0.3 kg N ha⁻¹; and continuously grown maize, fertilized with 195 kg N ha⁻¹, lost 0.1 kg N ha⁻¹.

This is not the only source of nitrogen reaching both land and water. Substantial and increasing amounts also arrive from the air. Elemental nitrogen is oxidized by lightning, in the course of burning plant materials, and in high-compression internal combustion engines, and biologically by the action of nitrogen-fixing soil bacteria. Urine from farm livestock releases ammonia, also a soluble compound. It has been found that in the mid-1970s much of Europe received 2–6 kg N ha⁻¹ yr⁻¹ and that some areas now receive 60 or more kg N ha⁻¹ yr⁻¹. This level of fertilization may be altering the composition of certain ecosystems, especially those established on nitrogen-poor soils (MOORE, 1995).

Plants have similar physiological requirements whether they grow on dry land or in water. If plant nutrients enter water, therefore, they will stimulate the growth of aquatic plants. Nitrate alone is not enough, of course. The full range of nutrients must be supplied and plant growth is limited by the availability of the nutrient in shortest supply (in water this is usually phosphorus); this is the ‘law of the minimum’ first stated in 1840 by the German chemist Justus von Liebig (1803–73). Other nutrients are less mobile than nitrate, so nitrate leaching has less effect on plant life than might be supposed.

Agricultural change apart, the movement of nutrients from the land and into water is an entirely natural process, an inevitable consequence of the drainage of rain water. As water moves downward through the soil to join the ground water, soluble soil compounds dissolve into and are carried by it. Were this not so, freshwater aquatic plant life would be severely restricted. Water draining into surface waters, such as rivers and lakes, also carries fine particulate matter that is deposited as sediment when the power of the stream falls below a certain threshold. Fast-flowing streams rapidly remove material that enters them and accumulations occur only in slow-moving rivers and still water. It is there, and only there, that sedimentation and eutrophication may cause difficulties.

Eutrophication leads to the proliferation of aquatic plants, especially algae, and cyanobacteria, organisms that derive nutrients directly from the water, rather than through roots attached to a substrate. A eutrophic lake or pond can usually be recognized by its surface covering of green algae. The life cycles of such organisms are short and as they die their remains sink and are decomposed by aerobic bacteria, whose populations increase in proportion to the food supply available to them. The bacteria obtain the oxygen they need from that dissolved in the water, and under eutrophic conditions the amount they remove exceeds the amount being introduced, so the water is depleted of dissolved oxygen. A common measure of water pollution is its ‘biochemical oxygen demand’ (BOD), calculated from the reduction in the amount of dissolved oxygen in a water sample incubated in darkness for 5 days at a constant 20°C; it is also a measure of bacterial activity.

If the water body is used for water abstraction, angling, or navigation, eutrophication is likely to reduce its value. The cost of treating water to bring it to potable standard will increase, navigation may be impeded by plants, and preferred species of fish may disappear. At high densities, some algae and cyanobacteria produce potent toxins. The alga *Prymnesium parvum* is highly toxic to fish, and toxins produced by such cyanobacteria as *Microcystis*, *Aphanizomenon*, and *Anabaena* attack the liver and may be neurotoxic. In 1989 there were outbreaks of toxic cyanobacteria in some British lakes and a number of dogs died after swimming in them and ingesting their water. Not surprisingly, eutrophication also brings about marked changes in the populations of aquatic organisms. The water supports fewer plant and animal species, but more individuals, the water becomes more turbid because of the large amount of organic matter suspended in it, the water becomes increasingly anoxic, and the rate of sedimentation increases.

A eutrophic lake is an old lake, and eutrophication is an ageing process. When it first forms, a lake typically supports little plant life, but fish such as trout, which feed on insects caught at the surface, may thrive. Its water is clear and well oxygenated, but very deficient in nutrients. There is little or no sediment at the bottom and plants grow beside it, but well clear of the water. A lake in this condition is said to be ‘oligotrophic’ (the Greek *oligos* means ‘small’ and *trophe* ‘nourishment').

Rivers flowing into the lake bring nutrient and particulate matter, and in time the lake becomes ‘mesotrophic’ (Greek *mesos*, ‘middle’). Its water is still clear enough for light to penetrate deeply, so algae flourish, but without proliferating uncontrollably because they are grazed by a diverse population of invertebrate and vertebrate animals, including fish. Sediment is accumulating on the bottom. This provides anchorage and nutrient for rooted plants, which now extend from the banks and into the lake margins, colonization by plants that have to reach the air being limited only by the depth of water. The accumulation of sediment also raises the bottom, so the lake has become shallower. In a eutrophic lake (Greek *eu-*, ‘well’) the sediment is deep and the lake shallow. Plants rooted in the sediment extend far from the banks. The three drawings in Figure 3.4 illustrate this life cycle.

Life cycles, which paradoxically are linear so far as individuals are concerned, end in death, and the life cycle of a lake is no exception. It is the fate of all lakes and ponds eventually to become dry land or, if they occupy low-lying ground where the water table is at or very close to the surface, a bog, marsh, or fen. Accumulating sediment makes the water shallower, but its colonization by plants also removes water, by transpiration. Once plants are established across the whole area of a lake, its demise is fairly rapid. Aquatic plants give way step by step to land plants that can tolerate waterlogging around their roots, and then these are replaced by true dryland or wetland plants. As the sediment dries and becomes soil, it is the acidity of the soil that determines whether the lake evolves into lime-loving grassland and, over much of north-western Europe, from there to scrub followed by woodland and forest, or to acid-loving heath. Figure 3.5 illustrates this development.

Such eutrophication is natural, but the life span of a lake should be measured in thousands of years. Artificial eutrophication, caused by discharging sewage and other wastes into lakes, shortens it greatly. Untreated human sewage may have a BOD of 300 mg litre⁻¹, paper-pulp effluent 25000 mg litre⁻¹, and silage effluent 50000 mg litre⁻¹. Deoxygenation is by far the commonest type of freshwater pollution. Bacteria decomposing the faeces from one human use 115 g of oxygen a day; this is enough oxygen to saturate 10000 litres of water (MELLANBY, 1992, p. 88). Halting natural eutrophication may be undesirable, even if it is practicable, but artificial eutrophication should be prevented or, if it is too late for prevention, cured.

It can best be remedied, of course, by finding alternative means of waste disposal or at least by reducing the nutrient content of the discharges, especially of phosphates, which are the limiting nutrient in most waters. This can be done by reducing the phosphate content of detergents, which

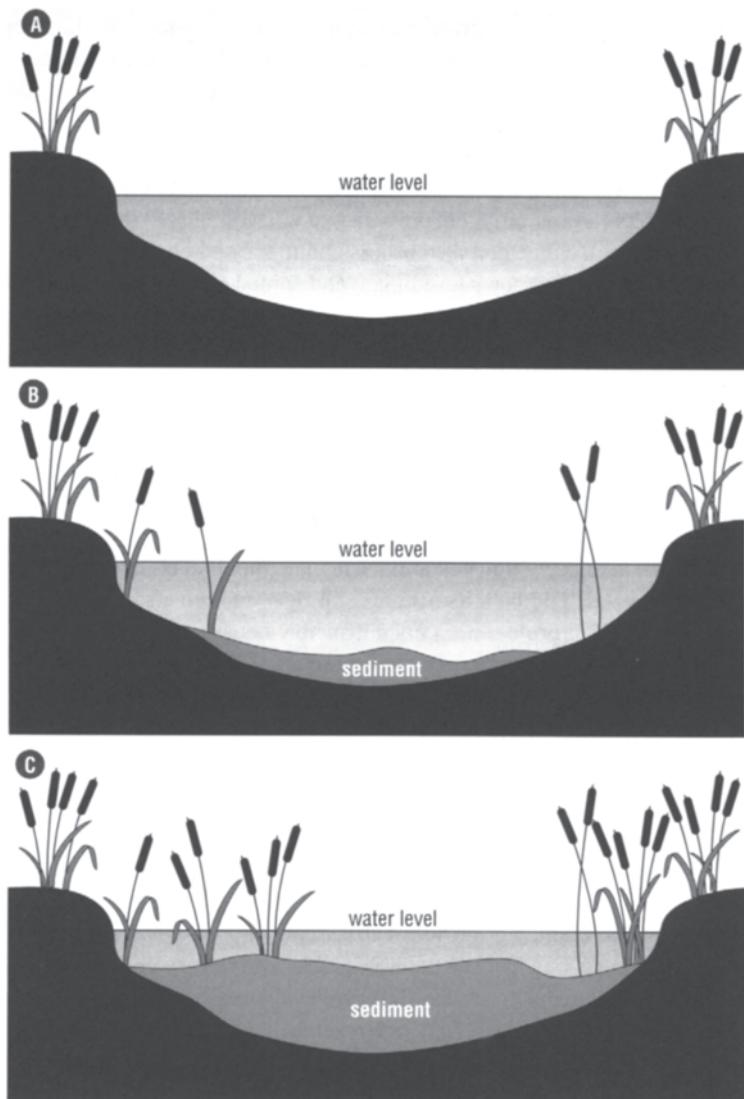


Figure 3.4 The life cycle of a lake. A, Oligotrophic. Little bottom sediment; water nutrient-poor; plants grow on banks only. B, Mesotrophic. Mud accumulating on the bottom; plants rooted in mud extending into the lake; moderate nutrient supply. C, Eutrophic. Deep bottom sediment; plants rooted in mud far into the lake; water very rich in nutrients; depth of lake decreasing owing to accumulation of sediment and evapotranspiration

are the principal source, or by stripping the phosphate from sewage before it is discharged. This is possible, with 90–95 per cent efficiency (MASON, 1991, p. 131), but there have been cases of a reduction in phosphate input being followed by the release of phosphate from sediment by mechanisms which are not well understood. In extreme cases it may be feasible to remove the sediment itself by dredging. Where land drainage is the main source of sediment and nutrient, reducing soil erosion may be effective. If oligotrophic water is available, using it to recharge a eutrophic lake may bring benefits. Beyond such measures as these, remediation usually involves manipulating the plant and animal populations. Obviously, no two water bodies are precisely similar and remedial measures must be appropriate to the particular conditions encountered.

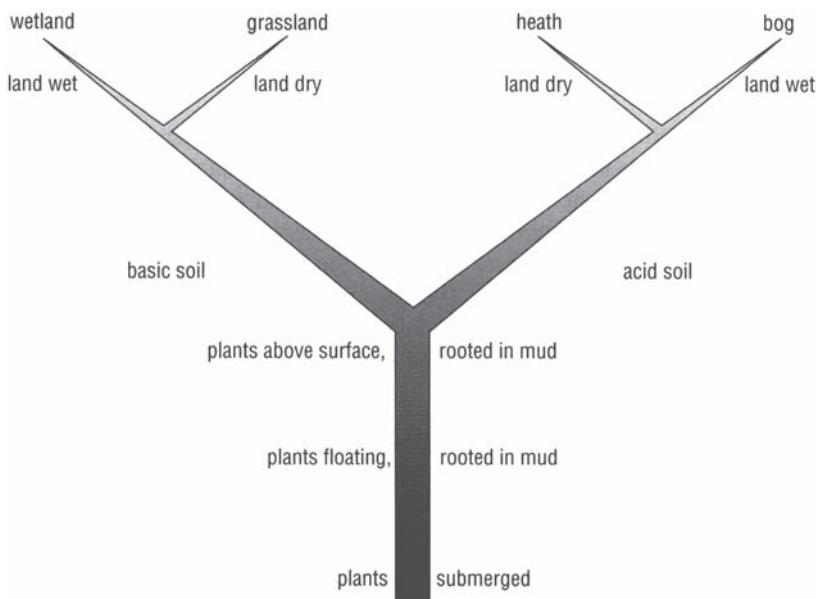


Figure 3.5 Evolution of a lake into dry land, marsh, or bog

It is easy to over-dramatize the problems of eutrophication. They are confined to still or slow-moving waters, which limits their extent. Nevertheless, remediation is often necessary, because the affected water body represents a valuable resource, and it is always complicated and expensive. Prevention being better than cure, control of discharges into surface waters, introduced primarily to improve the quality of river water that is not liable to eutrophication, will nevertheless reduce eutrophication in lakes fed by the improved rivers. The principal cause of river pollution is identical to that which produces artificial eutrophication.

24 Salt water, brackish water, and desalination

Water is a scarce resource in many parts of the world. Even in regions where rainfall is usually adequate, periodic droughts can bring shortages, and restrictions on water use are fairly common in Britain, despite its generally moist, maritime climate. These restrictions have never been so severe as to direct serious attention to alternative sources of supply, however, except on some offshore islands, such as the Isles of Scilly, in the Western Approaches off Land's End, where a desalination plant has been proposed.

Since almost all the water on Earth is in the oceans, sea water is the most obvious place to seek supplies and, after all, nowhere on the Isles of Scilly is more than a mile or so from the sea. The disadvantage of sea water, of course, is its salt content. Industrial plants located in coastal areas can use sea water directly for cooling, which is why many British nuclear power plants are located at the coast, but sea water is useless for agricultural or domestic purposes. The cells of living organisms are contained within membranes that are partially permeable, allowing water molecules to pass, but blocking the passage of larger molecules, in a process known as 'osmosis'. If a partially permeable membrane separates two solutions of different concentrations, an osmotic pressure will act across the membrane, forcing water molecules to pass from the weaker to the stronger solution until the

concentrations equalize. When cells are exposed to sea water, its salt concentration is higher than the concentration inside the cell, and water moves out of the cell. Salt water thus has a dehydrating effect and its salts must be removed before land-dwelling plants or animals can use it.

This is expensive, and there is another source of fresh water: the polar icecaps. The idea may sound absurd, but probably it would be technologically and economically feasible to tow large icebergs into low latitudes, moor them close to the shore, and 'mine' them for fresh water. An iceberg would begin to melt as it entered warm water, but the rate of melting would be low enough to ensure the survival of the great bulk of the ice and the loss would be acceptable. Clearly, the resource is vast and possibly self-renewing. There is a major disadvantage, however. Because the communities needing the water are located far inland, but the iceberg is at the coast, water must still be transported over a long distance. Combined with the cost of towing, this would probably make the operation prohibitively expensive.

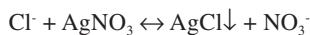
'Iceberg mining' has not yet been attempted and neither has a rival scheme, suggested by Walter Rickel, Governor of Alaska, to construct a submarine pipeline to carry water 3220 km to California from the headwaters of Alaskan rivers. The scheme was considered, but rejected because of its estimated \$100 billion cost (REINHOLD, 1992).

Desalination (www.ce.vt.edu/enviro2/wtprimer/desalt/desalt/html), on the other hand, is used widely in the Near and Middle East. It is also used in the United States. For some years the Office of Saline Water, of the Department of the Interior, has maintained a demonstration desalination plant at Freeport, Texas, and there is a large plant in Arizona. More recently, water shortages in California led to the construction of a plant at Catalina yielding 580280 litres of fresh water a day, and plants are also to be built at Santa Barbara and Morro Bay (REINHOLD, 1992).

The purpose of desalination is the removal of salts from sea water, but not all sea water is equally saline. Together, temperature and salinity determine the relative densities of different water bodies, which form water masses analogous to air masses. Plotted on a graph, seawater masses can be identified by their position along a temperature-salinity (T-S) curve. Salinity is conventionally reported in parts per thousand (per mille). In the centre of the North Atlantic, for example, the T-S curve ranges from 8°C and 35.1 per mille to 19°C and 36.7 per mille; around Antarctica the seawater temperature is 2–7°C and salinity 34.1–34.6 per mille (HARVEY, 1976, pp. 61–63). Elsewhere, salinity may be markedly higher or lower. The Mediterranean loses more water by evaporation than it receives from inflowing rivers and precipitation; it also loses water at depth and gains inflowing water near the surface through the Straits of Gibraltar. This regime results in a salinity higher than that of the Atlantic, ranging from about 37.0 per mille near Gibraltar to about 39 per mille at the eastern end. The Black Sea, in contrast, has an average salinity of about 19.0 per mille, the Caspian 12.86 per mille, and the Red Sea 41.0 per mille (DAJOZ, 1975, pp. 126–128). Variable though the salinity of sea water is, it remains true that all sea water is too salty to drink: fresh water has a salinity of less than 0.3 per mille.

Water that is neither fresh nor sea water is known as 'brackish' and its salinity is even more variable. Oligohaline water is only slightly more saline than fresh water, with a salinity of 0.5–5.0 per mille; mesohaline water has 5.0–16.0 per mille; polyhaline water has 16.0–40.0 per mille; and saline water has more than 40.0 per mille. Water in the Great Salt Lake has a salinity of 170 per mille and that in the Dead Sea 230 per mille. Yet all these are 'brackish' waters.

Salinity is measured by titrating a sample of water with silver nitrate until all the chloride ions have been precipitated, and adding potassium chromate, which reacts with silver nitrate when all the chloride has been precipitated, forming potassium chromate, which is red. The reaction is:



In other words, what is being measured is *chlorinity*.

Regardless of its salinity, or chlorinity, the composition of sea water is fairly constant (Table 3.1). Some of the ‘other’ (listed at the foot of Table 3.1) is of commercial importance, actually or potentially. It contains about 3 parts per million of uranium, for example, and about 0.003 per cent of all water, including sea water, is deuterium oxide, or ‘heavy water’, used as a moderator in the Candu (Canadian deuterium-uranium) fission reactors and, in years to come, as a fuel in fusion reactors. Interpreted ironically, the percentage composition of sea water is shown in Table 3.2. Fresh water has a much more variable composition, but one dominated by carbonates (79.9 per cent) and sulphates (13.2 per cent), with chlorides contributing only 6.9 per cent.

Removing the dissolved salts from sea water leaves a highly concentrated brine. Those salts for which industrial markets can be found can be extracted and sold. Common salt, metallic magnesium, magnesium compounds, and bromine are obtained in this way. Indeed, nearly 30 per cent of the world supply of salt is obtained by evaporating sea water. In this process, calcium sulphate and calcium carbonate precipitate first; when they have been removed the brine is moved to another pond, where salt crystallizes. The remaining brine, called ‘bitterns’, is removed, fresh, concentrated brine is added, and this is repeated until the layer of crystalline salt is thick enough to be harvested. Bromine can then be extracted from the bitterns. Where no market for by-products can be found, however, disposal of the brine is difficult, and for every 30000 litres of fresh water produced by desalination, 1 tonne of salts remains.

Water may be separated from its dissolved salts by distillation, freezing, electrolysis, or reverse osmosis. Distillation is the most widely used method. In low latitudes, the Sun may supply enough energy to evaporate sea water. The evaporate is then condensed and after several cycles of evaporation and condensation the water is sufficiently pure to be fed into the public supply. More usually, however, energy must be provided. Several distillation methods are used. Figure 3.6 illustrates multistage flash evaporation, which is one of the most efficient. Incoming sea water is heated under pressure, to prevent it from boiling, then released into a chamber where pressure is lower. It boils instantly ('flash boiling') and the vapour rises, to condense on the pipe carrying cold, incoming sea water. The latent heat of condensation warms the incoming water, reducing the amount of heating required. The condensate is collected and removed and the remaining brine fed to the next chamber where the process is repeated.

Ice contains little salt and so freezing sea water purifies it. In this technique, the sea water is chilled almost to its freezing temperature, then either sprayed into a partly evacuated chamber or mixed with a volatile hydrocarbon, such as butane, and poured into a chamber. The low pressure, or high volatility of the hydrocarbon, causes immediate evaporation of the hydrocarbon or some of the water and the chilling caused by the latent heat of evaporation causes some of the remaining water to freeze. The slurry of ice and brine is then pumped into another chamber, fresh water is added to separate ice from brine, and the fresh water is removed.

Table 3.1 **Composition of sea water**

Constituent	%
Sodium chloride	77.8
Magnesium chloride	9.7
Magnesium sulphate	5.7
Calcium sulphate	3.7
Potassium chloride	1.7
Calcium carbonate	0.3
Other	1.1

Table 3.2 **Ions in sea water**

ions	%
Chloride	55.04
Sodium	30.61
Sulphate	7.68
Magnesium	3.69
Calcium	1.16
Potassium	1.10
Bicarbonate	0.41
Bromide	0.19
Other	0.12

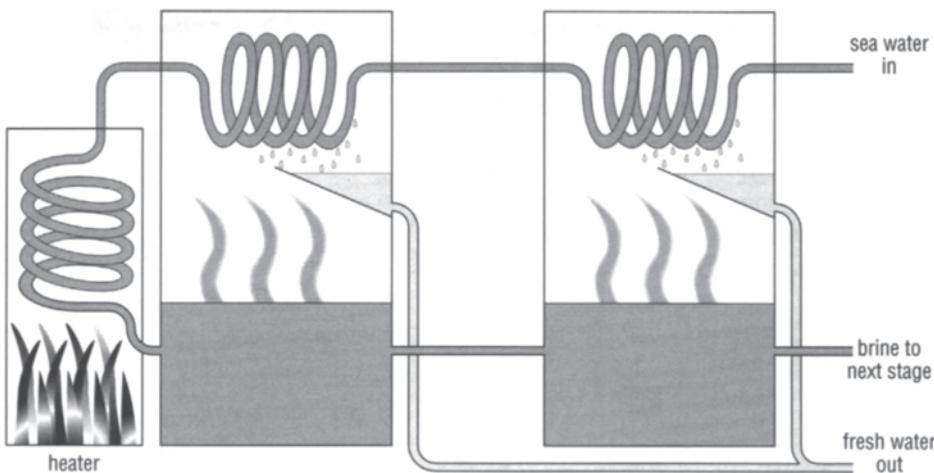


Figure 3.6 Multistage flash evaporation

Osmosis

If two solutions of different strengths are separated by a membrane that allows molecules of the solvent to pass, but not those of the solute (the dissolved substance), solvent molecules will cross the membrane from the weaker to the stronger solution until the two are of equal strength. The membrane separating them is called 'differentially permeable' if it allows water molecules to pass but slows the passage of larger molecules or prevents some of them, or 'semi-permeable' if it is completely permeable to molecules of solvent and completely impermeable to those of the solute. Cell membranes are differentially permeable. Membranes that allow the passage of some but not all molecules are now often described as 'partially permeable'.

The passage of water through a membrane requires energy. Pure water is considered to possess zero energy and a solution to have a negative energy value. Osmosis occurs when there is an energy difference between two solutions and the energy involved, known as the 'osmotic pressure' or 'water potential', can be measured.

In reverse osmosis sufficient pressure is applied to a solution to overcome the water potential and force water molecules to cross a semi-permeable membrane from the higher to lower concentration. The pressure required is about 25×10^5 Pa (25 times ordinary sea-level atmospheric pressure).

Electrolytic desalination involves pumping sea water into a chamber containing electrodes. Some ions are attracted to the positive electrode (anode), others to the negative electrode (cathode) and partly purified water is extracted from the middle.

As its name suggests, reverse osmosis is based on a natural process. A partially permeable membrane separates fresh from sea water and the pressure of the sea water is increased. The high pressure

required makes reverse osmosis difficult to apply on a large scale, but advances made in recent years have reduced the energy needed to below that required for distillation and the technique is becoming commercially attractive.

In years to come, rising demand for fresh water will lead to greater reliance on desalination. At present, the high energy requirement makes all industrial-scale desalination technologies too expensive for many of the less developed countries, where the increased demand will be felt most acutely, but this situation could change. More efficient techniques for exploiting solar energy might reduce costs in low latitudes, and in high latitudes waste heat from coastal industrial plants, especially nuclear power stations, might be used to the same end.

As the production of fresh water by desalination grows, however, so will the amount of highly concentrated brine for which no economic use can be found. It would be as well to develop satisfactory means for its disposal before proceeding rapidly along this path.

25 Irrigation, waterlogging, and salinization

Deprived of water, before long any plant other than a cactus or other succulent will begin to look very sick indeed. Its leaves will become flaccid and if it lacks a woody stem the entire plant will grow limp and collapse. It will wilt. The condition may be temporary, the plant recovering when its access to water is restored, but if it continues for too long the wilt will be permanent and the plant will die.

Plants need water to give rigidity to their cells, but water stress also produces other, more subtle effects. The stressed plant will spend more time with its stomata closed. These are the pores, each opened and closed by the expansion and contraction of a pair of guard cells, through which gases are exchanged and from which water evaporates. Keeping stomata closed reduces water loss, but a reduction in the rate of gas exchange implies a reduction in the rate of photosynthesis. The plant will grow more slowly and will be smaller than it would otherwise be, and growth is inhibited before the plant is so short of water that it wilts visibly. When an adequate amount of water becomes available to a formerly stressed plant it will increase its production of foliage, but in the case of a crop plant its final weight will never be greater than that of an unstressed plant and usually it will be smaller.

Water shortage is an obvious problem facing farmers in semi-arid climates, or in climate types with pronounced wet and dry seasons, such as that of the Mediterranean. Less obviously, it can also reduce agricultural production where rainfall is distributed fairly evenly through the year. The monthly extent of water surplus or deficit can be calculated by comparing the amount of rainfall with the amount of water lost by evaporation and transpiration from grass supplied with abundant water. Such calculations show that in central England a water deficit may occur during the summer and autumn, from June to October, when evaporation exceeds precipitation (WINTER, 1974, p. 7). If water is provided in addition to that received as rainfall, field experiments at the National Vegetable Research Station in England have shown that crop yields increase dramatically: those of maincrop potatoes rose from 37 t ha^{-1} to 50 t ha^{-1} , an increase of 13 t ha^{-1} , and those of cabbage from 41 t ha^{-1} to 59 t ha^{-1} , an increase of 18 t ha^{-1} . For every 25 mm of irrigation per hectare, yields of main-crop potato increased by 3 t and those of cabbage by 18 t (WINTER, 1974, p. 117).

Irrigation is clearly beneficial, even in much of Britain, but this is hardly news. Farmers were irrigating their crops seven thousand years ago in Mesopotamia and irrigation techniques were developed independently in China, Mexico, and Peru. In some countries unirrigated agriculture would be impossible; all farm land is irrigated in Egypt, for example. In the world as a whole, about 15 per

cent of all farmland is irrigated, ranging from 6 per cent in Africa and South America to 31 per cent in Asia. Between 1970 and 1990 this area increased by more than a third, from 168 million ha to 228 million ha, most of the increase being in developing countries, and the output from irrigated land is more than double that from unirrigated land; one-third of the world's food is grown on irrigated land (TOLBA AND EL-KHOLY, 1992, p. 290).

Water for irrigation is often provided by damming rivers to fill reservoirs, the flow of water from the dams also generating electrical power, but large dams can produce adverse environmental effects. Their reservoirs flood large areas, destroying existing plant and animal communities and often displacing many people, and silt carried from upstream tends to accumulate, gradually filling the reservoir. Where rivers formerly flooded land downstream at a certain time of the year, the silt deposit containing plant nutrients is lost to farmers, who must buy fertilizer to replace it. In seismically active regions, large dams may also be linked to increases in the number of earthquakes. An earthquake exceeding magnitude 5 on the Richter scale occurred while the first large dam in the world, the Hoover Dam on the Colorado River, was being filled in 1936 and there was another of comparable magnitude in 1939. There have also been earthquakes greater than magnitude 5 associated with the Koyna Dam, India (1967), Kremasta Dam, Greece (1966), Hsinfengkiang Dam, China (1962), and Marathon Dam, Greece (1938), each of them accompanied by foreshocks and aftershocks (GOUDIE, 1986, pp. 243–244).

Land can be irrigated simply by flooding it and allowing the water to sink into the ground. A somewhat more sophisticated method is to dig parallel furrows down the slope of a field and fill them with water from a ditch or pipe across the upper edge of the field. A more familiar technique involves the use of sprinklers. These are versatile, in that they can be moved to where they are most needed and the amount of water they deliver can be controlled closely. In some places, irrigation is supplied by subsurface pipes.

Environmentalists used to be fond of saying ‘everything has to go somewhere’. This is as true of water as of anything else and water supply is only one side of the water management equation: water must also be removed. In some places, wet ground can be rendered cultivable only by making it drier; in others, irrigation must be accompanied by improved drainage.

Land drainage is a farming practice probably as ancient as irrigation. On sloping ground, a ditch along the upper boundary of a field, at right angles to the direction of slope, will collect water draining from higher land before it flows into the field. A network of communicating ditches can then carry the surplus water to the nearest stream.

On level ground, or where the construction of ditches is insufficient, drains may be laid below ground. The simplest technique is to install ‘mole’ drains, so called because the implement that makes them tunnels through the soil like a mole. The ‘mole’ itself is a metal cylinder fixed to the lower end of a bar, buried to the desired depth, and then towed through the soil. Figure 3.7 illustrates the device and shows that it makes a hole parallel to the surface. In most soils the hole will remain open for some years before the operation needs repeating. More permanent drains are made from short lengths of perforated piping laid end to end by a machine that digs the trench into which it lays them, then buries them as it passes. In both cases the drains feed into a stream or system of ditches. The land area drained is proportional to the depth of the drain, so it is a simple matter to plan a drainage system that will serve a whole field without leaving wet patches.

It is easy to see why farmers find it desirable to remove surplus water from wet ground. The need for a drainage system to accompany irrigation is less self-evident, but lack of drainage on irrigated land is a major cause of soil degradation.

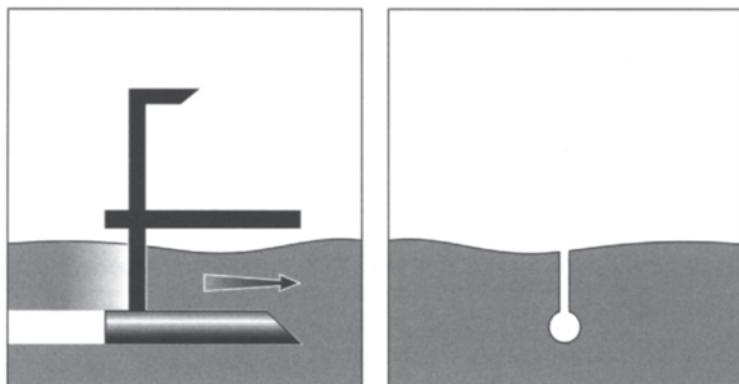


Figure 3.7 Mole drainage. Left, mole plough; right, cross-sectional view of mole drain

If more water is abstracted from an aquifer than flows in to replenish it, the amount of water available will gradually diminish. In coastal regions, such over-exploitation of the resource brings an additional hazard. Beneath the sea bed, the sediment is permanently saturated with salt water. The salt water moves inland, beneath the freshwater aquifer, with a boundary of brackish water separating the two water bodies. As the freshwater aquifer is depleted, this boundary moves further inland and closer to the surface, allowing salt water to penetrate the soil. As Figure 3.8 shows, a point can be reached at which water abstracted for irrigation starts to become brackish and the more that is abstracted the saltier it is. Since most crop plants are very intolerant of salt, the effect can be to sterilize the affected land. It is a problem in many coastal areas, but especially serious in low-lying islands, such as coral atolls (TOLBA AND EL-KHOLY, 1992, p. 117).

This form of contamination is known as 'salinization' (or in the USA as 'salination'). Salt water intrusion can occur only in coastal areas, but salinization quite unconnected to the proximity of sea water affects regions far inland. According to the UN Environmental Programme (UNEP), 7 million ha is affected in China, 20 million ha in India, 3.2 million ha in Pakistan and the Near East, and 5.2 million ha in the United States. Parts of southern Europe also suffer from this problem (TOLBA AND EL-KHOLY, 1992, p. 290). It arises because of the way water moves through soil.

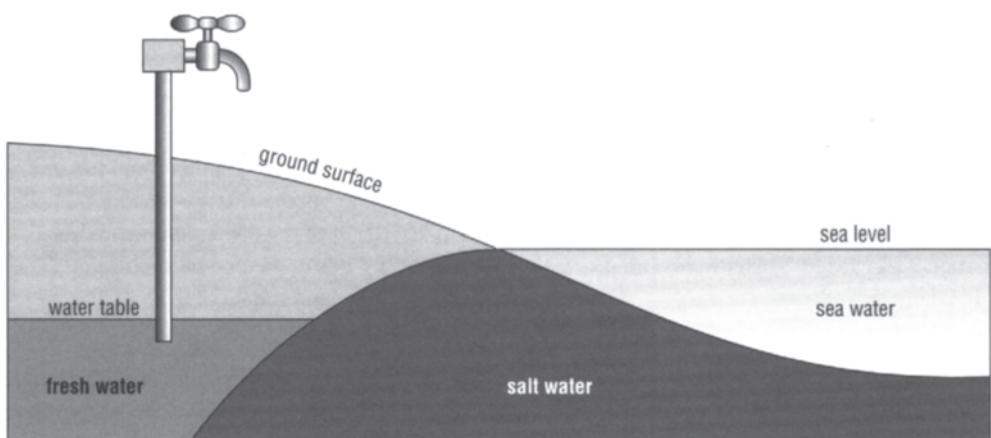


Figure 3.8 Salt water intrusion into a freshwater aquifer

Some of the rain falling on the ground sinks vertically through the soil, as ‘gravitational water’, until it reaches the ground water, the region where the soil is saturated, the upper boundary of the saturated region being the water table. Above the water table, particles comprising the unsaturated soil are coated by a very thin film of ‘adhesion water’ held by attraction between water molecules and the electrically charged surfaces of soil particles. Even the driest dust is usually coated with adhesion water. This film is covered by an outer film of ‘cohesion water’, held by the attraction of hydrogen bonds between water molecules themselves.

Water molecules at the bottom of a pot of water, or adjacent to the impermeable material underlying the ground water, are subject to a pressure equal to the weight of water above them. The higher they are in the pot, the less pressure bears down on them, until, at the surface, the pressure is zero. Any water in tiny but connected spaces above the surface will be under even less (i.e. negative) pressure: it will be under tension, a force pulling it upward rather than downward. Molecules will be easily attracted by the adhesive charge on soil particles and further molecules will join them because of the cohesive attraction of the molecules already in place. This is capillary attraction. It has little effect on adhesion water, which moves very little, but cohesion water is less tightly bound and can move. Under soil moisture tension, it moves to coat dry soil particles (becoming adhesion water) and to equalize the thickness of the layer of cohesion water throughout the soil. Very slowly, the water will rise through the unsaturated layer, and a very small suction by a plant root hair will be sufficient to dislodge cohesion water and move it into the plant (FOTH AND TURK, 1972, pp. 64–74).

Water rises through both plants and soil, evaporates, and is replaced by more water rising through the capillary pore spaces in the soil. Water vapour is almost pure H₂O, and any substances dissolved in the liquid are precipitated as it evaporates. Soil water is far from pure. Salts dissolve into it as it moves through the soil and some soils contain quite large amounts of soluble salts. Irrigation water itself is seldom pure; farmers do not irrigate their land with water fit for human consumption. The water they use commonly contains between 750 g m⁻³ and 1.5 kg m⁻³ of dissolved salts (FOTH AND TURK, 1972, p. 407). These may be left as evaporates near the soil surface, deposited from water that evaporated before soaking into the soil or from water that descended gravitationally and then rose again by capillary attraction. Gradually, the salinity of the upper soil increases until plants begin to suffer, the most salt-intolerant first, but eventually most crop species.

Salinization most commonly occurs in arid or semi-arid climates, where the rate of evaporation is high, but it is under these conditions that irrigation is most urgently needed and where it may bring its greatest benefits. The risk may be avoided by installing adequate drainage to remove surplus water before it can evaporate and by controlling the dissolved-salt content of irrigation water, especially on saline soils.

Should it occur, the remedy is slow, difficult, and expensive. The first area ever to have been irrigated, in the Tigris and Euphrates valleys, suffered from salinization and to this day much of it remains barren because its reclamation would be too costly. Fresh water, containing little or no dissolved salts, must be used to flush the salts from the soil and into a drainage system that will remove them, and it may be necessary to take care in disposing of the salt-laden water. If salinization was caused by salt-water intrusion, the freshwater aquifer must also be recharged. The old adage still applies, of course: the water used to clean saline land must come from and go to somewhere.

Over-zealous irrigation on poorly drained land can lead to a quite different problem. If more water is added to the soil than can evaporate or be transpired by plants, the water table will rise. It may do so for some time before the consequences become apparent, but eventually soil around the roots of crop plants will be saturated and, being saturated, airless. No water may be visible lying on the surface, but nevertheless the land is waterlogged and crop yields will fall dramatically. In this case the remedy

is simpler. Adequate drainage must be installed and irrigation suspended until the water table has been lowered.

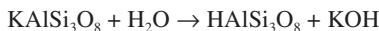
As the demand for food intensifies it is likely that the total area of irrigated land will increase. Some say it may double between about 1990 and the early part of the next century (PIEL, 1992, p. 216). In Asia, where such an increase is likely to be concentrated, this will allow two or even three crops to be grown each year on land that presently produces only one. The advantages will prove enduring, however, only if irrigation schemes are planned with care to avoid the hazards attendant on them.

26 Soil formation, ageing, and taxonomy

From the moment it is exposed at the surface, rock is subjected to persistent physical attack. Water fills small fissures and when it freezes it expands, exerting a pressure of up to 146 kg cm^{-2} , which is sufficient to split the toughest rock (DONAHUE *ET AL.*, 1958, p. 28). In summer, the rock warms during the day and cools again at night, expanding as it warms and contracting as it cools, but it is heated unevenly. The surface is heated more strongly than rock beneath the surface; some parts of the surface are exposed directly to sunlight, others are in shade. As a consequence, some parts of the rock expand and contract more than others. This, too, causes the rock to break. Often flakes are loosened or detached from the surface, a process called ‘exfoliation’. Detached particles then grind against one another as they are moved by gravity, wind, or water. This breaks them into still smaller pieces.

The smaller any physical object, the greater its surface area in relation to its volume: a sphere with a diameter of 4 units has a surface area of 50 units² and volume of 33.5 units³, giving an area:volume ratio of 1:0.7; if the diameter is 2, the surface area is 12.5 units², volume 4.2 units³ and ratio 1:0.3. As the rock particles grow smaller, therefore, the total surface area exposed to attack increases. Still vulnerable to abrasion, they are now subject to chemical attack.

This takes several forms. Some of the chemical compounds of which they are composed may be soluble in water; wetting dissolves and drainage removes them. Other compounds may react chemically with water. The process is called ‘hydrolysis’ and can convert insoluble compounds to more soluble ones. Orthoclase feldspar (KAlSi_3O_8), for example, a common constituent of igneous rocks, hydrolyses to a partly soluble clay (HAlSi_3O_8) and very soluble potassium hydroxide (KOH) by the reaction:



Hydration is the process in which compounds combine with water, but do not react chemically with it. The addition of water to a compound’s molecules makes them bigger and softer and so increases their vulnerability to breakage. Oxidation also increases the size and softness of many mineral molecules and may also alter their electrical charge in ways that make them react more readily with water or weak acids. Reduction, which occurs where oxygen is in short supply, also alters the electrical charge on molecules and may reduce their size.

Compounds may also react with carbonic acid (H_2CO_3), formed when carbon dioxide dissolves in water. This reaction, called ‘carbonation’, forms soluble bicarbonates. Barely soluble calcium carbonate (CaCO_3), for example, becomes highly soluble calcium bicarbonate ($\text{Ca}(\text{HCO}_3)_2$).

Physical and chemical processes thus combine to alter radically the structure and chemical composition of surface rock. How long it takes for solid rock to be converted into a layer of small mineral particles depends on the character of the original rock and the extent of its exposure; in arid climates

it proceeds more slowly than in moist ones, for example. Yet the process is remorseless. At widely varying speeds it dismantles mountains.

It does not proceed far before living organisms accelerate it: respiration and the decomposition of plant remains are the main source of the carbon dioxide engaged in subsurface carbonation. The chemical changes release compounds useful to organisms in soluble forms they can absorb, and their metabolic wastes and dead cells add to the stock of reactive compounds as well as providing sustenance to still more organisms. Bacteria are usually the first to arrive, forming colonies in sheltered cracks, invisible to the naked eye. Lichens often follow, composite organisms comprising a fungus and alga or cyanobacterium. The fungus obtains water and mineral nutrients from the rock, the alga or cyanobacterium supplies carbohydrates that it photosynthesizes and oxygen as a by-product of photosynthesis. Each partner supplies the other and the fungus protects them both from drying out and provides firm attachment to the rock, which it grips tightly with filaments that grow into the tiniest crevices. This remarkable partnership allows lichens to flourish where no plant could survive.

Organic material, derived from wastes and the decay of dead cells, accumulates beneath the lichen, mixing with the mineral particles and accelerating chemical reactions. This mixture is better at absorbing and holding water, and in time there is enough of it to provide anchorage and nutrients for plants. Mosses may arrive and small herbs may root themselves in the deeper cracks.

As the layer of mixed organic and mineral material thickens, some of it begins to be washed to deeper levels, a few centimetres below the surface. The material is starting to form two distinct layers: an upper layer from which soluble compounds and particles are being washed (the technical term is 'leached') and a lower layer in which they are accumulating. This is the first stage in the formation of soil.

From this point, vegetation becomes part of the developing soil and contributes greatly to its formation. Plant roots penetrate the material and when they decay leave channels that assist aeration and drainage. Dead plant material contributes fresh organic matter to the surface, which decays to release compounds that drain into the soil. In detail, however, this process can vary widely over a small area, in large part because of the efficiency with which the soil drains and the depth of the water table below the surface. If the soil is derived from similar mineral particles all the way down a slope, a hydrologic sequence may occur (CRUICKSHANK, 1972, pp. 47–48), illustrated in

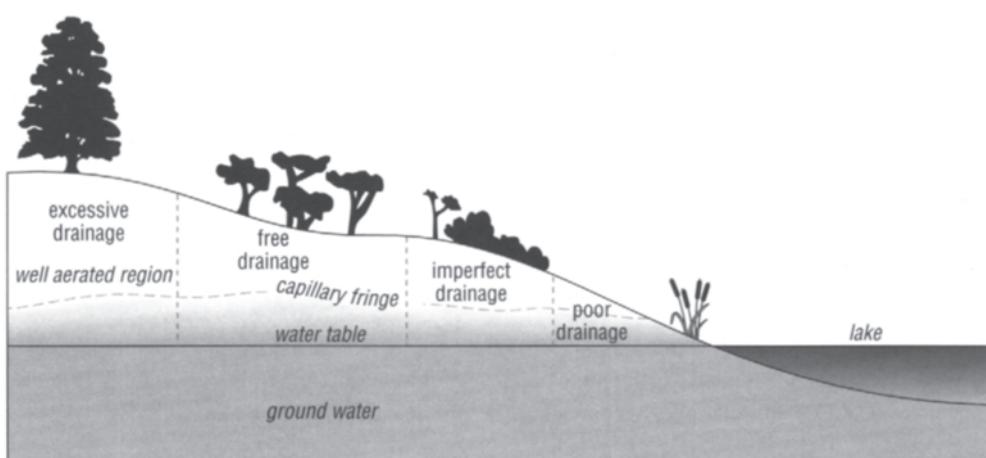


Figure 3.9 **Soil drainage**

Figure 3.9. Where drainage is excessive, the soil will be generally dry, favouring trees that can root to considerable depth. As the distance narrows between the surface and water table, the region hospitable to plant roots becomes shallower, and the plants smaller. Most of the organisms engaged in the decomposition of plant material require oxygen for respiration, so the decreasing depth of the aerated zone is accompanied by a slowing of the rate of decomposition until, where the soil is waterlogged, partly decomposed material may form acid peat.

The stage of this process that involves purely physical and chemical mechanisms constitutes weathering; as living organisms become the predominant agents it is called ‘pedogenesis’. Plants growing at the surface penetrate the soil with their roots and supply a topmost layer of dead organic material, called ‘litter’. This provides sustenance for a diverse population of animals, complete with their predators and parasites, fungi, and bacteria. These break down the material, which enters the soil proper, much of it carried below by earthworms, where it feeds another population. Compounds released by decomposition dissolve in water draining through the soil and are carried to a lower level, where they accumulate. At the base of this layer, the ‘subsoil’, rocks and mineral particles, detached from the underlying rock, are being weathered, and below this layer lies the bedrock itself.

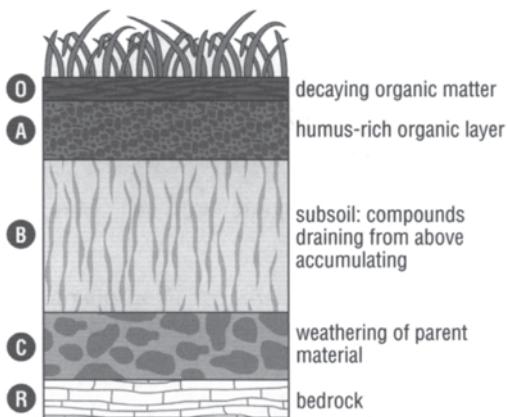


Figure 3.10 **Profile of a typical fertile soil**

If a vertical section, called a ‘profile’, is cut through the soil from surface to bedrock, it may reveal this structure as layers, called ‘horizons’, clearly differentiated by their colour and texture. Figure 3.10 shows the principal horizons, but in any particular soil there may be more or fewer and in some soils horizons are not easily distinguished at all. Conventionally, the horizons are identified by letters: O for the surface layer of organic matter; A for the surface horizons; B for the accumulation layer; C for the weathering layer; and R for the bedrock. The horizons are classified further by the addition of numbers: A2 is a mineral horizon somewhat darkened by the presence of organic matter; A3 is a transition zone between the A

and B horizons. Letters are then added in subscript to denote particular characteristics: ca means the soil contains calcium and magnesium carbonates; g (for ‘gleying’) means the soil is poorly aerated and frequently waterlogged; m means the soil is strongly cemented together, like a soft rock.

Soils vary according to the rock from which they are derived. This affects the size of their mineral grains, ranging from coarse sand (600–2000 µm) to silt (2–60 µm) and clay (less than 2 µm), and their chemical characteristics. Soils derived from granites, for example, develop slowly, are usually sandy, and contain relatively few plant nutrients; those developed from limestones are usually fine-grained and relatively rich in plant nutrients.

Once formed, soils begin to age. The rate at which they do so depends principally on the climate and vegetation. Desert soils age slowly, and so do those in polar regions, but in the humid tropics soils age much more quickly as luxuriant plant growth extracts nutrients and returns them for decomposition into soluble forms, which are leached rapidly by the abundant water. It is possible, therefore, to describe as ‘young’, ‘mature’, or ‘ancient’ soils that may have been in existence for the same length of time.

We obtain our food from soil, we erect buildings of varying weight upon it, and we use clay taken from it as construction material that may or may not be fired to make bricks. Clearly it is of great

importance to us and if we are to use it the more we know about it the better. It is so variable that we cannot be satisfied in calling it simply ‘the’ soil. It must be classified.

There have been many attempts at soil classification, the first in classical times, but it was not until the latter part of the last century that a school of Russian scientists at St Petersburg, led by Vasily Vasilievich Dokuchaev (1840–1903), proposed a theory of pedogenesis on which a formal classification could be based. It is because of this Russian origin that many soil types have Russian names, such as ‘podzol’ and ‘chernozem’. The Russian work laid the foundation for what is now known as ‘soil taxonomy’, but work has continued ever since.

The system most widely used at present was devised by the US Department of Agriculture. This divides all soils into 11 orders (www.explorer.it/aip/keytax/content.html). The orders are divided further into sub-orders, great groups, families, and soil series. The orders, with brief descriptions, are listed in the box.

The 11 soil orders of the US soil taxonomy

Alfisols	Soils of climates with 510–1270 mm annual rainfall; most develop under forests; clay accumulates in the B horizon.
Andisols	Volcanic soils, deep and light in texture; contain iron and aluminium compounds. (This order is sometimes omitted.)
Aridisols	Desert soils with accumulations of lime or gypsum; often with salt layers; little organic matter.
Entisols	Little or no horizon development; often found in recent flood plains, under recent volcanic ash, as wind-blown sand.
Histosols	Organic soils; found in bogs and swamps.
Inceptisols	Young soils; horizons starting to develop; often wet conditions.
Mollisols	Very dark soils; upper layers rich in organic matter; form mainly under grassland.
Oxisols	Deeply weathered soils; acid; low fertility; contain clays of iron and aluminium oxides.
Spodosols	Sandy soils found in forests, mainly coniferous; organic matter, iron and aluminium oxides accumulated in B horizon; strongly acid.
Ultisols	Deeply weathered tropical and subtropical soils; strongly acid; clay accumulated in B horizon.
Vertisols	Clay soils that swell when wet; develop in climates with pronounced wet and dry seasons; deep cracks appear when dry.

These names and descriptions are reasonably straightforward, but the system becomes much more abstruse below the level of orders. Suborders have such names as ‘Psammements’, ‘Boralfs’, and ‘Usterts’; the great groups include ‘Haplargids’, ‘Haplorthods’, and ‘Pellusterts’; and among the subgroups are ‘Aquic Paleudults’, ‘Typic Medisaprists’, and ‘Typic Torrox’.

The classification may be powerful, but there are attractions in calling Mollisols 'prairie soils' (or chernozems), Histosols 'peat' or 'muck', and, given the widespread environmentalist concern over the degradation of some tropical soils, calling Oxisols 'lateritic soils', which are the names by which they used to be known. There is also a Canadian classification system that divides soils into two orders, Brunisolic comprising 4 Great Groups and Chernozemic with 3 Great Groups and a total of 42 Subgroups (burgundy.uwaterloo.ca/bio1446/chap2bsm.htm).

27 Transport by water and wind

Once it has developed, soil does not necessarily remain in the place where it formed. It can be transported, sometimes over very long distances, by wind and water, and over short distances by gravity.

Occasionally, rain in northern Europe leaves everything exposed to it coated by a thin layer of reddish dust. It is Saharan dust, lifted from the desert, carried some 2500 km by air movements, then washed to the ground. Some of that dust remains to form a very minor constituent of European soil. Fine soil particles lofted by wind in the North American Great Plains during the dust-bowl years of the 1930s fell as dust in New York and discoloured the Atlantic for hundreds of kilometres from the American coast.

The Dust Bowl

An area of about 390000 km² in south-western Kansas, south-eastern Colorado, north-eastern New Mexico, and parts of Oklahoma and Texas that was originally prairie. The climate is semi-arid and prolonged droughts are common. After 1918, US grain prices rose steeply, encouraging farmers to plough the prairie, and for several years they produced satisfactory yields. A drought began in 1933 and lasted until 1939, being especially severe in 1934 and 1935. It was in those two years that topsoil, exposed by ploughing and reduced by aridity to a fine dust, blew away, the lighter particles forming clouds 8 km high. Farmers, most of whom were already poor, were ruined and thousands of them migrated to seek other work.

In 1935 the US Department of Agriculture founded the Soil Conservation Service to promote sound soil conservation practices throughout the country, and with the return of the rains some farming resumed, although much of the area was returned to grass. Drought has continued to afflict the region at intervals of about 20–22 years.

Where aeolian (wind-blown) deposits accumulate they make a soil known as 'loess'. Obviously, the smaller the particles the further the air will carry them, and so the material is graded, becoming finer with increasing distance from its source. Loess soils are extensive, covering much of the central United States, where in places they are several metres thick (HUNT, 1972, p. 138). There are also extensive deposits in Argentina, various parts of Europe, and in China; in the northern and eastern highlands they are believed to be 300 m thick (DONAHUE *ET AL.*, 1958, pp. 24–25), with the underlying rock projecting through it locally as hill and mountain ranges.

Although deserts are a source of wind-blown dust, modern dust storms make only a minor contribution to such deposits; most are ancient, dating from past ice ages. When glaciers thawed, they released

meltwater that flowed as rivers and flooded low-lying areas. The waters carried suspended particles, which settled as mud. Then, as the temperature fell, the meltwater flow ceased, the flooded areas dried, and wind blew the dust far away from the valley bottoms. The process was repeated at intervals, some of them long enough for new soil to begin to form above the loess before being buried beneath a later deposit. Loess soils are usually yellowish in colour and when young they are rich in mineral plant nutrients and calcium. This makes them inherently fertile, although their later history may exhaust that fertility. They are fine-grained, and rivers cutting through loess commonly have very steep, almost vertical, banks. Much of south-east England was once blanketed by such loessic soils.

As you will know if you have walked over a dry, sandy beach in a strong wind, sand grains also blow. Being much larger than the silt-sized particles that comprise loess, they are not carried far, but repeated lifting and dropping by a prevailing wind can transport them a considerable distance. Where they accumulate, dunes form, sometimes with a characteristic shape from which the wind direction can be determined. Crescent-shaped (called ‘barchan’) dunes, for example, are aligned across the direction of the prevailing wind, the convex side facing the wind behind a long ‘tail’ of sand being blown little by little up the tail to the top, where the dune collapses on the sheltered side to produce the face; strong, steady winds erode valley-like troughs and linear dunes; dunes forming straight ridges may be parallel (seif dunes) or at right angles to the wind direction (aklé dunes); and changing winds produce star dunes, of radiating ridges.

Blown sand is unsuitable for cultivation, but it can be stabilized if hardy plants become established on it. In temperate climates marram grass (*Ammophila arenaria*) is often used for initial colonization in coastal areas. Its underground stems (rhizomes) form networks that help hold the sand in place and allow other plants to obtain purchase. As a diverse plant community develops the marram grass is unable to compete and disappears. Then soil will form above the sand, burying it. Unstabilized sand dunes migrate slowly and can bury fertile land downwind.

Water is a much more powerful agent of transport than wind. Gravel and small stones cannot be carried far by even the strongest wind, but a river can move them long distances. A soil that is mineralogically unrelated to the bedrock beneath it and that contains sand mixed with stones of various sizes has been placed in its present position by water. It may be a marine deposit, formed on the bed of a sea that has long since disappeared. If the particles are sorted by size into layers, the deposit is more likely to be lacustrine, marking the location of a former lake. Many lacustrine deposits contain a high proportion of clay particles; if the material is more than 50 per cent clay it will be almost impermeable to water and thus prone to waterlogging or flooding.

Glaciers also transport material, but they seldom carry it very far. Their action is mainly to mix the soil already formed beneath them and to transport large pieces of rock that become frozen into the ice. Then, when the glacier retreats by melting at its lower end, these stones join the mixed soil to make ‘till’ (which was formerly known as ‘boulder clay’). Till deposits cover substantial areas in Europe and North America. Although glaciers rarely carried the till more than 10 km, large stones entrained in the ice were sometimes carried much further and deposited as ‘erratics’. Their orientation and that of stones in the till itself can be used to determine the direction in which the ice was moving. Because glaciers filled or made broad, flat-bottomed valleys, glacial till often occurs in gently rolling ‘till plains’. Material pushed to the sides of a glacier and ahead of its upward-curving front, or ‘snout’, was left as a moraine, now visible as ridges or hills that are often too rocky to be easily cultivable.

Permafrost occurs in the vicinity of glaciers and ice sheets. This is ground where the temperature below the surface remains below freezing throughout the year. In summer the surface layers may thaw, and if the soil is on a slope greater than about 2°, the resulting mud will flow, carrying with it

stones released by frost-shattering from adjacent rock. These stones, most of very irregular shape, will tend to become aligned with their long axes indicating the direction of slope and parallel to the surface. At the base of the slope the flowing material will collect as 'head', often forming a thick deposit. The sliding of material down a slope over ground that is frozen a little way below the surface is called 'gelifluction' or 'congelifluction'. It is the cold-climate variety of solifluction, the downslope creep of material lubricated by water.

Glacial meltwaters often flowed with great force. They carried huge volumes of water, sometimes through confined spaces so they were under pressure, and they carried particles of all sizes, including quite large stones. As the flow slowed, the heaviest particles were deposited first and such 'outwash' material is usually coarse-grained and sorted by grain size into beds. The finer-grained material travelled further and settled as mud on land that dried as the water flow ceased or on the bottom of glacial lakes, many of which have since disappeared. Glacial lacustrine deposits consist mainly of clay of little agricultural value, but in some places used for brick-making.

Soils derived from river-borne sediment are called 'alluvial' and occur on land that is, or was, repeatedly flooded. Many rivers burst their banks occasionally, but for the flooding to affect soil formation they must do so often, and this is most likely if their waters are periodically augmented by drainage following very heavy rain or seasonally by the melting of deep snow. Charged with a greater volume of water to transport, they flow faster and this gives them the energy to transport more material. When the river overflows its banks, the water escapes to the sides and the pressure on it is greatly reduced. It loses energy and deposits its load, the heaviest particles first. These may collect close to the point where the overflow occurred, after many floodings forming a distinctive raised bank, called a 'levee'. This may become covered with soil, and small particles will be trapped within it, but basically it is made from gravel and larger stones, and water drains through it freely. Silt and clay are precipitated beyond the levee. They pack tightly together and drain poorly, but are rich in plant nutrients and the flood plain they produce as they fill natural depressions and make a level surface is usually very fertile.

Downstream, where the river, fed by many tributaries, flows as a wide stream across land with a very low gradient, it may form meanders. A system of meanders may also create a flood plain, but by a quite different mechanism that does not require the river to burst its banks. On the outside of each meander bend, the stream flows against the bank. This increases its turbulence and also its speed, since it has slightly further to travel, and material is drawn into the water from the bank, eroding it. Some of this material may flow across the stream, near the bed, where it enters water flowing against the inside bank. Here the flow is slower and calmer. The river loses energy and deposits some of its load, extending the bank into the channel. Figure 3.11 illustrates the process.

At the same time, the entire meander tends to migrate in a downstream direction and since this movement affects all the meanders, the entire system migrates downstream. The land behind a migrating meander is covered by river-bed material, so an alluvial plain forms that is the same width as the widest meander. Meanders migrate slowly, so the fertile flood plain can be used, although the ground may be wet for most of the time.

Till and alluvium, as well as loess, are commonly buried beneath soils developed since they were deposited, so they may not be visible at the surface. They will be detectable, however, because they form the parent material for the soil covering them. This soil may not resemble its parent, but will betray its origin by being unrelated to the underlying bedrock.

A section through the soil will reveal its character, of course, and a river may cut a suitable section. As Figure 3.12 shows, a stream channel quickly penetrates the surface material to expose the

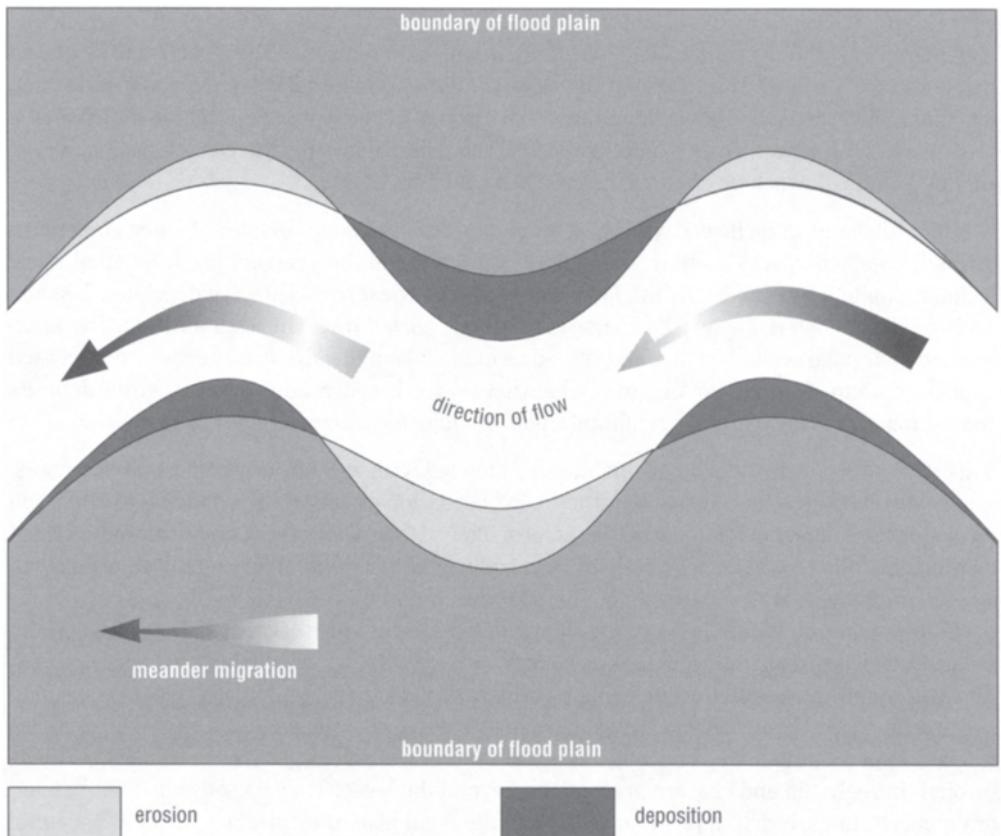


Figure 3.11 Flood plain development from meander system

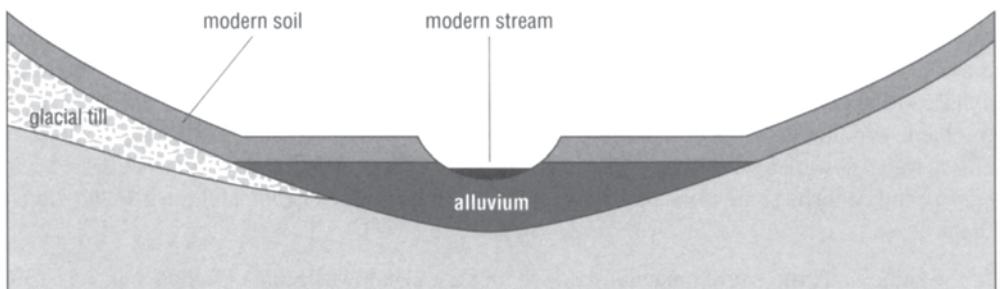


Figure 3.12 Modern soil developed over flood plain alluvium and glacial till

underlying deposit, in this case suggesting that a much larger river once flowed, probably fed by glacial meltwater which also left a layer of till.

Soil is formed by processes at or very close to the surface. Once formed, it is affected by other processes, some of which tend to transport it to new locations. The consequences of these processes may be serious, catastrophic even, for humans living where they occur. In the central plains of North America farmers from the east cleared away the natural prairie grassland and ploughed the land to grow wheat in a climate drier and more prone to drought than that to which they were

accustomed. Following a series of dry years, in 1934 and 1935 the crops failed and soil reduced to little more than powder was carried away by the wind, creating the Dust Bowl (ALLABY, 1998, pp. 76–82). This brought tragedy to countless American farming families in the 1930s, most of whom were already impoverished as a result of the economic depression. After that experience it was recognized that the land was unsuitable for arable farming and much of it has been returned to grassland. To take another example, unexpected floods destroy crops and livelihoods, and may cause the deaths of livestock and people. Yet the disasters may also bring benefits to those farming the fertile loess or alluvial soils.

Such events occur naturally. Clearly, the soils of the Dust Bowl were unsuitable for the type of farming practised on them. By removing the natural vegetation cover and cultivating the soil, farmers reduced the soil to a condition in which it would blow away, but the drought was a natural occurrence. Human activities can interfere more directly. When a river is dammed, for example, silt will accumulate behind the dam. This progressively reduces the volume of water held in the artificial lake, but it also interrupts the natural sedimentation process further downstream. On the flood plain or delta, farmers may depend on the seasonal floods for the silt they bring, rich in the plant nutrients that have drained into it the entire length of the river's course. Deprived of their 'natural fertilizer', farmers may be forced to buy factory-made fertilizer, which in many cases they can ill afford, and the farming techniques they have developed may be inappropriate for alluvial soils that are not regularly replenished. Soil structure and fertility may deteriorate. Similarly, the clearing of vegetation from upper slopes may increase the transport of sediment to lower levels, polluting rivers.

28 Soil, climate, and land use

Climate is by far the most important factor controlling the development of soils (pedogenesis) from their parent materials, but it is not the only one. Pedogenesis is also affected by the type of vegetation (which is also climatically governed), the activities of animals including humans, the parent material, and topography. Most of the chemical reactions by which the mineral constituents of rock are modified do not commence until the temperature rises above 10°C, and their reaction rates double for every further 10°C rise above that. It follows that, provided water is available, soils develop and age faster in warm than in cold climates and it is in warm climates that biological organisms are more active.

Some of the early schemes for classifying soils were based on climate, grouping soils into cold, cool-temperate, subtropical, and tropical zones, within which many of the names were descriptive. In the cold zones, for example, are tundra and mountain meadow soils, and in the cool-temperate zones prairie soils (CRUICKSHANK, 1972, pp. 155–166). This led to the concept of 'zonal soils', typical of the zones in which they occur. The zones might also contain soils formed under the influence of some local factor and therefore atypical. These were called 'intrazonal soils'. Soils that had not developed at all might be found in any climatic regime, and were called 'azonal'.

Such classifications were based on examinations of only the A and C soil horizons, the B horizon being considered merely transitional between those above and below it. In time, soil scientists came to realize that zonal schemes were classifying not so much the soils as the environments in which they form. Modern soil taxonomy is based on the soils themselves, defined in terms of more than 20 surface and subsurface 'diagnostic horizons', called 'epipedons' and 'endopedons' respectively. An anthropic epipedon, for example, is a surface horizon formed where people have lived for a long time or have grown irrigated crops; cultivation over many years might lead to the formation of an agric endopedon just beneath the ploughing depth, where clay and organic matter have collected.

Although soils are now classified according to their composition, their surface horizons are formed biologically, by the mixing of organic and mineral material. Since natural vegetation is usually typical of the climate in which it grows, it is impossible to dissociate pedogenesis and climate.

As Figure 3.13 shows, the relationship between plants and soil is intimate. Beneath a conifer forest, there is usually a deep layer (A0 horizon) of organic material, mainly needles. This decomposes only slowly, partly because conifer needles have a thick, waxy, outer layer that is not easily broken. This is an adaptation to climate: conifers grow in climates with a pronounced dry season or a long winter when water is frozen and so unavailable. The dry or cold season also reduces the rate of decomposition. The A1 horizon, dark and rich in humus (decayed organic material), is thin and the somewhat thicker A2 horizon is very pale, because its humus has leached into the B horizon. This soil is classified in the order Spodosols.

Broad-leaved forests produce a much thinner A0 horizon, because the more delicate leaves, shed in the autumn, decompose fairly quickly during the mild, wet winter. The resulting humus forms a deep, dark, A1 horizon, a thinner, leached, A2 horizon, and a deep B horizon, where plant nutrients accumulate well within the reach of tree roots. This is an Alfisol.

Mollisols, found beneath temperate grassland, also have a thin A0 horizon, because grass produces a dense but shallow mat of roots. Organic matter decomposes rapidly. The humus-rich A1 horizon is deep and the leached A2 horizon correspondingly thin, with a deep B horizon, where nutrients accumulate.

It is the Aridisols that provide dramatic contrast. Developed under desert conditions, they support almost no vegetation and consequently have no surface litter at all. Because there is no A0 horizon, there can be no A1 horizon either, because no humus is being produced. Occasional rains produce weathering, and soluble compounds are leached into a deep B horizon, beneath which there is a further horizon where calcium carbonate accumulates.

The combination of climate and vegetation has further effects on soils. Spodosols are prone to ‘podzolization’. The slow decay of organic matter releases acids that drain downward, removing carbonates as they do so, and making the whole of the A horizon acid. In extreme cases it may be so acid as to cause the leaching of clays and their accumulation as a hard, impermeable layer (a ‘hardpan’) in the B horizon. In permafrost regions, the winter freezing of the surface layer causes it to expand, compressing the soil beneath against the underlying permafrost. This inter-

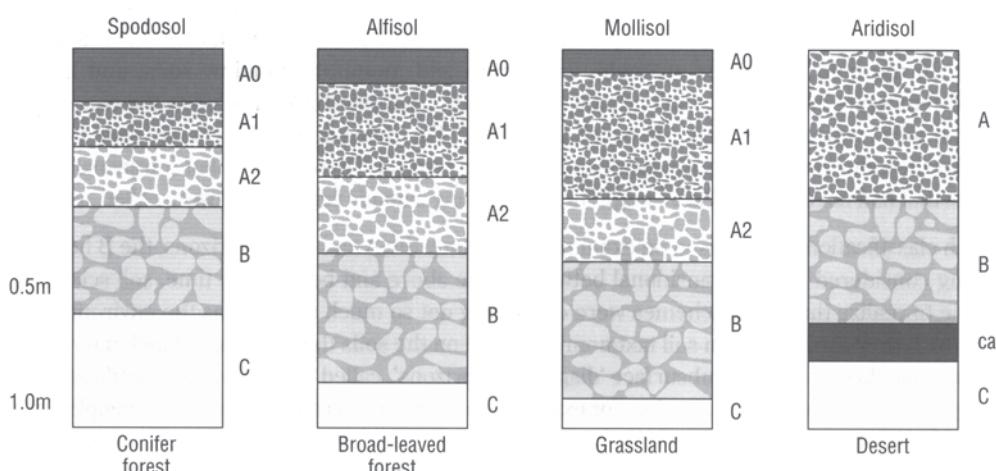


Figure 3.13 Profiles of four soils, with the vegetation associated with them

mediate layer, called ‘gley’, is wet, sticky, and blue because of the reduced iron compounds it contains, and compression forces it upward through cracks and mixes it with the overlying material in the process of ‘gleying’ (or in the United States ‘gleyzation’).

Oxisols, the predominant soils of the humid tropical lowlands, are the deepest of all soils. In many places the total soil depth, from surface to bedrock, may be 10 m and all the lower horizons are very thick. Scale apart, they are similar to Spodosols, except that their surface horizons may be badly eroded, so the B horizon is close to the surface, and they contain very little humus, because of the rapidity with which organic matter decomposes under the humid tropical climate, its nutrients being reabsorbed by plants. Almost all plant nutrients are contained within the living vegetation itself and soluble compounds have been leached from the soil, leaving it acid and inherently infertile. Clays, mainly comprising kaolinite (an aluminium-silica mineral, $\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$) and ferric and aluminium oxides and hydroxides, may accumulate near the top of the B horizon, cemented together to form nodules or more extensive layers of ‘laterite’. Laterite is extremely hard and impermeable and, in soils prone to it, laterization, the process by which it forms, is sometimes accelerated by clearing vegetation and leaving the ground exposed to heavy rain and thus to increased leaching.

Given the close association between climate and pedogenesis, it is not surprising to discover that the global distribution of soil orders broadly conforms to climatic zones. As Figure 3.14 illustrates, Oxisols are found in the humid tropics, Alfisols in temperate regions, Mollisols in the prairies, pampas, and steppes, and Spodosols in a belt around northern America and Eurasia.

Simply because an apparently deep, dark soil occurs in a climate favourable for agriculture, however, it does not necessarily follow that the land will sustain farming. Farmed soils are ‘domesticated’ by years of careful management and are markedly different from the ‘virgin’ soils that preceded them. Early farmers settled on the most promising land and when, after a few seasons, their crop yields began to decline, they moved elsewhere and started again. There is only a limited store of plant nutrients in any soil and it is depleted by the removal of crops, which reduces the amount available for recycling. Fertilizers and lime (to restore leached calcium) replenish the store, but if they are unobtainable (or unknown) farmers may have no alternative but to adopt some form of shifting cultivation. This still remains a common type of farming in many parts of the tropics.

Soil fertility is not determined by the amount of plant nutrients contained within reach of plant roots, because presence and proximity do not guarantee access. The roots must be able to absorb the nutrients they require and their ability to do so depends on the chemical characteristics of the soil.

Humus and silicate clays consist of masses of microscopic soil particles, each about 2 μm across. They are called ‘colloids’, because they change between a gel-like consistency and liquid according to the chemical environment around them. Soil colloids have negatively charged surface sites onto which cations (positive ions) can be adsorbed. Humus has many more of them than clay, which in turn has far more than sand. The commonest soil cations are calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^+), potassium (K^+), and aluminium (Al^{3+}), but they are changing constantly as one replaces another, often in the order $\text{Al} \rightarrow \text{Ca} \rightarrow \text{Mg} \rightarrow \text{K} \rightarrow \text{Na}$. Fertilizers add cations, such as ammonium (NH_4^+), and lime adds calcium.

Anions (negatively charged ions) are also exchanged, but to a much lesser extent. Several important plant nutrients commonly occur as anions, including sulphate (SO_4^{2-}), nitrate (NO_3^-), phosphate (H_2PO_4^- or HPO_4^{2-}), and molybdate (MoO_4^{2-}). These are not held at cation-exchange sites, but dissolve in soil water and are absorbed by plants directly from solution.

While the colloid remains saturated with exchangeable cations it retains its structure, but as these are replaced by hydrogen (H^+) the structure weakens. The soil becomes more acid (a measure

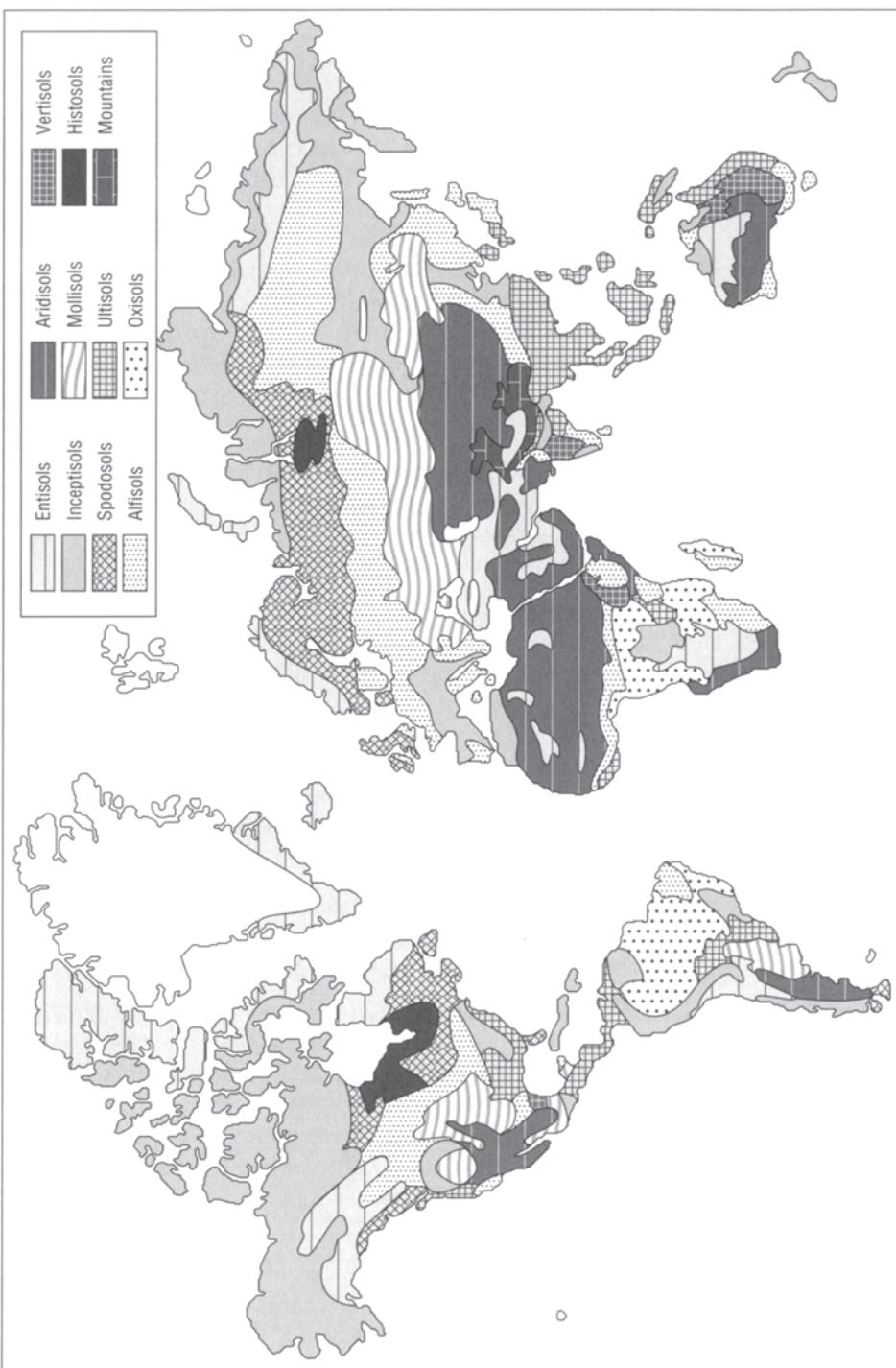


Figure 3.14 World distribution of soil orders

of the hydrogen ion concentration) and when it is nearly saturated with hydrogen the colloid breaks down into its constituent compounds, which move downward through the soil.

The amount of exchangeable cations in a unit weight of dry soil is known as the ‘cation exchange capacity’ (CEC) of that soil and it provides an important measure of soil fertility. It is measured in milliequivalents (me) per 100 g (a milliequivalent being that quantity chemically equal to 1 mg of hydrogen). Sandy soils typically have a CEC of 1–5 me 100 g⁻¹, loams 5–15 me per 100 g, and clays more than 30 me per 100 g. The CEC for a typical Mollisol ranges from just over 24 me per 100 g at the surface to 25.7 me per 100 g in the lower part of the B horizon. The proportion of the cation-exchange sites occupied by calcium, magnesium, potassium, and sodium ions is known as the ‘percentage base saturation’; the lower its value the more acid the soil. Most crop plants prefer a base saturation of 80 per cent or more, producing a neutral soil (pH 6.0 or higher), but some grow best in more acid soils, with a lower percentage base saturation (FOTH AND TURK, 1972, pp. 171–175).

Those cations that are plant nutrients are absorbed by roots while dissolved in soil water. Some are replaced from other exchange sites in the soil colloid, but these are generally insufficient to replace all of them, and cation-rich fertilizers must replenish the store, the frequency and amount of fertilizer application depending on the CEC of the soil. In other words, soil fertility can be expressed as CEC and sandy soils (low CEC) generally need more fertilizer more often than clay soils (high CEC).

Cation exchange also makes soils with a high CEC excellent purifying filters for water percolating through them. Positively charged pollutants, such as lead (Pb^{2+}) and cadmium (Cd^{2+}), are quickly adsorbed to exchange sites and thus immobilized, so water contaminated with them is purified by the time it reaches the ground water.

It is technologically possible to grow any plant anywhere in the world by supplying it with the environment it requires. Construct and heat a glasshouse, supply it with an appropriate soil, illuminate it artificially to achieve a suitable light intensity and length of day, and tropical crops could be raised in Greenland. It is possible, but hardly sensible when tropical crops can be produced so much more easily and cheaply in the tropics. Soils appropriate to particular plants are most likely to develop where the climate also suits those plants or their close relatives, but there are dangers. Soil fertility is usually reduced by cropping and must be replenished, and some soils, especially in the tropics, are much less fertile than the luxuriant vegetation they naturally sustain makes them appear.

29 Soil erosion and its control

Central Belgium has large areas of loess soils and it is estimated (KUPCHELLA AND HYLAND, 1986, p. 447) that each year these are losing between 10 and 25 tonnes of soil from each hectare by erosion. The United States is losing about 18 t ha⁻¹ yr⁻¹ and the Yellow River basin in China 100 t ha⁻¹ yr⁻¹. In 1939, wind-blown soil from Texas fell in Iowa, 800 km away. It landed on top of snow, so the soil could be collected, weighed, and analysed. The amount deposited was 450 kg ha⁻¹ and the deposit contained more than three times more organic matter and nitrogen than the soil from which it had been removed (HUDSON, 1971, p. 257). Clearly, soil erosion is a serious problem affecting cultivated land in almost every part of the world. As was mentioned earlier (Section 2) there is reason to hope that the introduction of modern farming methods may reduce the rate of soil erosion dramatically. Using herbicides to clear the ground of weeds as an alternative to ploughing is the most promising method. This makes it possible to reduce tillage to a minimum and to sow seeds directly into land covered with, and protected by, dead weeds. At the same time, increasing the productivity of the best land reduces the need to cultivate the poorer, more erosion-prone land.

In fact, soil erosion is an entirely natural process. Unconsolidated surface material is transported by wind and water from the moment it is exposed, whether the land is cultivated or not, and this transport has been continuing throughout the history of our planet. Many sedimentary rocks are made from eroded soil, after all, but the Earth is still blanketed by soil, which is constantly being formed. Erosion need not become a cause for concern until its rate exceeds the rate of soil formation; after that threshold is crossed, soil is actually being lost.

Under natural conditions, soil probably forms at an average rate of about 8 mm per century. Ploughing the land aerates the soil and increases the rate of leaching. This accelerates soil formation to perhaps 80 mm per century (HUDSON, 1971, p. 36). The planned reclamation of spoil heaps from mining can produce several centimetres of organically enriched surface soil within 5 years, but soil formation is considered to commence when substances leached from the A horizons start accumulating in the B horizons and the soil begins to acquire its layered structure. This takes much longer.

On cultivated land, therefore, soil is ordinarily forming at a rate of about $2 \text{ t ha}^{-1} \text{ yr}^{-1}$, which is the weight represented by the 0.8 mm forming each year. A loss greater than this indicates a net loss by erosion, but just how undesirable this is considered to be depends on the soil itself. If the soil is no more than a thin layer overlying the bedrock, for example, its erosion is much less acceptable than a similar rate of erosion from a very deep, fertile soil.

Whether a particular rate of erosion is acceptable is a matter of judgement, but the susceptibility of a field to erosion under different farming regimes can be estimated and the erosion rate predicted. The technique requires a number of factors to be calculated. The erosivity of the rainfall (R) is calculated by measuring the amount and type of rainfall, converting this into an index number, and reading R from a scale. The credibility of the soil (K) is a number representing the liability of a particular soil to erosion. The length factor of the field (L) is the ratio of the length of the field to the length of a standard field (of 22.6 m). The slope factor (S) is the ratio of the soil lost to the amount lost from a field with a 9 per cent gradient. The crop management factor (C) is the ratio of soil loss to that from a field under cultivated bare fallow. The conservation practice factor (P) is a ratio of soil loss to that from a field where no care is taken to prevent erosion. The amount of soil lost from that field each year (A) is then calculated from the universal soil-loss equation:

$$A = R \times K \times L \times S \times C \times P$$

Wind erosion can be spectacular and sometimes frightening, but water is probably the more important agent. Quite ordinary rain will initiate erosion, known as ‘splash erosion’, if it falls on bare soil. Raindrops fall at about 9 m s^{-1} , giving them a kinetic energy 13.6 times their own weight (DONAHUE *ET AL.*, 1958, pp. 323–325). This is sufficient to detach soil particles, in the case of fine sands and silt splashing them to a height of up to 60 cm and a distance of up to 1.5 m. Larger particles do not move so far and clay particles are held together by their strong cohesive attraction, but even these will travel some distance. Splashing does not remove soil directly, but moves particles around and the water carries them into every tiny opening. At the same time, the pounding of the rain packs the surface particles together. When the rain ceases the surface dries as a tough, impermeable crust sealing the soil beneath. The next time it rains, water is unable to drain vertically and must flow across the surface, carrying soil particles with it in a thin mud, down the slope and away.

The effects of splash erosion can sometimes be seen on banks beside roads or ditches, where tree roots or a large rock have sheltered the soil. These are partly exposed, because the soil has been washed away from around them. This is known as ‘pedestal erosion’.

Water flowing across the surface carries particles in suspension, which never touch the ground, particles that slide or roll over the surface, and particles that are repeatedly lifted and dropped. Some of the water finds its way into natural depressions and flows along them. This widens and deepens them, forming ‘rills’. Rills are small enough to be removed by ordinary cultivation, so they are temporary, but unless they are removed they may grow much larger, into ‘gullies’. These are more difficult to remove, because ordinary farm implements cannot cross them.

Routine cultivation destroys rills and ensures that gully erosion is rare on cultivated land, but on land that is never ploughed it can be serious. Tracks used by vehicles can turn into streams in wet weather, and from streams into badly eroded gullies. Paths much used by walkers suffer in the same way. A vicious spiral develops, in which the worn-down path becomes wet, walkers avoid the mud by walking to the side, and the gully widens as well as deepens. Some years ago in England the southern end of the Pennine Way national trail had to be rerouted, so severe was the erosion on Kinder Scout.

Two steps are involved in the removal of soil: the detachment of particles, by rain splash or wind; and their transport by water or wind. The remedy lies in minimizing both.

When rain falls on to vegetation, or wind blows across ground covered by vegetation, particle detachment is greatly reduced. The raindrops and wind are dispersed and their energy dissipated as leaves absorb the shock and rebound, like springs. For arable farmers this may be more difficult than it sounds; land is usually bare between the time of sowing and the emergence of the crop. Nevertheless, some sort of cover is valuable on vulnerable soil. In some places this is achieved by planting crops in alternate strips, such as grain and grass, or by leaving stubble lying on the surface after harvesting. In areas with a temperate climate, such as Britain, cereals are sown in autumn, as soon as possible after the completion of the previous harvest. This allows the seed to germinate and provide a vegetation cover through the winter, and a crop that starts growing rapidly in spring. It minimizes erosion, but is possible only where winter temperatures do not fall so low as to kill the young plants. In regions with a more extreme climate, cereals must be sown in spring and the soil must remain bare through the winter, although in this case erosion is reduced by the freezing of the surface or by a covering of snow.

Contour ploughing, in which the plough follows the contours of the land, produces parallel furrows oriented at right angles to the slope. Soil eroded from the plough ridges is trapped in lower furrows. In a field ploughed up and down the hill, with furrows parallel to the slope, soil can be swept downslope in furrows that quickly become rills.

Where slopes are steeper and the land is farmed intensively, terraces may be desirable. There are many types and Figure 3.15 shows cross-sections of two of the most common. Broad-base terraces are made by cutting trenches not quite at right angles to the slope and at intervals down it, using the excavated soil to build ridges on the downslope sides. Soil being washed down the hillside is held before it has travelled very far and the gently sloping trenches carry the water into drains or ditches. This avoids the situation in which soil erodes from the top of a field and accumulates at the bottom, making the soil quality uneven.

Bench terraces require more drastic engineering, because the hillside must be converted into a series of level strips, like a staircase, excavated soil being used to construct a bank along the downslope edge of each terrace. The technique is very effective, and is widely used in the tropics, but because the terraces must follow the contours they can have irregular shapes that make machine cultivation awkward.

Surplus water can be removed by ‘grass waterways’. These are wide strips running down the slope and sown to grass. In effect, they are controlled gullies and can follow the routes of actual

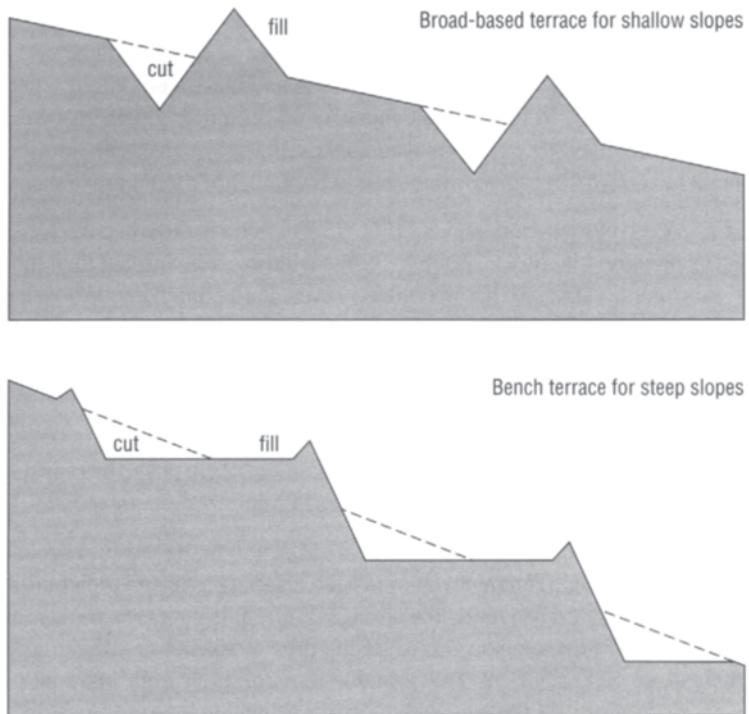


Figure 3.15 Two types of terracing for reducing runoff

gullies, but much more productively. During dry weather the gully is widened, soil added, and grass seed sown beneath a mulch held in place by netting. When the grass is established the waterway will continue to carry water, but will trap soil, and the grass itself can be grazed or cut for hay or silage.

Netting can also be laid over a bare surface after seed has been sown to achieve temporary erosion control. This holds the soil in place until plants have emerged to bind it more permanently. The technique is sometimes used on roadside verges and road-centre reservations, and the netting can be made from material that slowly decomposes and is incorporated into the soil.

Techniques for preventing wind erosion are designed to minimize particle detachment by reducing the speed and thus the energy of the wind. Maintaining a vegetative cover achieves this by creating a relatively calm ground-level microclimate. Strips of short stubble about 3 m wide will catch and hold almost all the soil lifted by the wind (FOTH AND TURK, 1972, p. 371).

Imagine the direction of the prevailing wind as a slope, as though it were blowing downhill, and a version of contour ploughing is also very effective. Where rows of crops are planted at right angles to the prevailing wind, each row shelters those downwind. Crops that are not grown in rows, such as cereals, can be protected by strips of row crops planted at intervals across the wind. As well as reducing erosion, this also helps prevent the soil from drying.

On a larger scale, trees and shrubs are often used as shelter belts or windbreaks. As Figure 3.16 shows, the wind is deflected over the top of the windbreak and its speed is reduced for a considerable distance downwind. The advantage of this is obvious, the disadvantage less so. Reducing wind speed also reduces the drying and chilling of the soil surface and can produce microclimatic conditions that vary markedly with distance from the windbreak. The resulting uneven growth and ripening of the crop can cause serious harvesting difficulties.

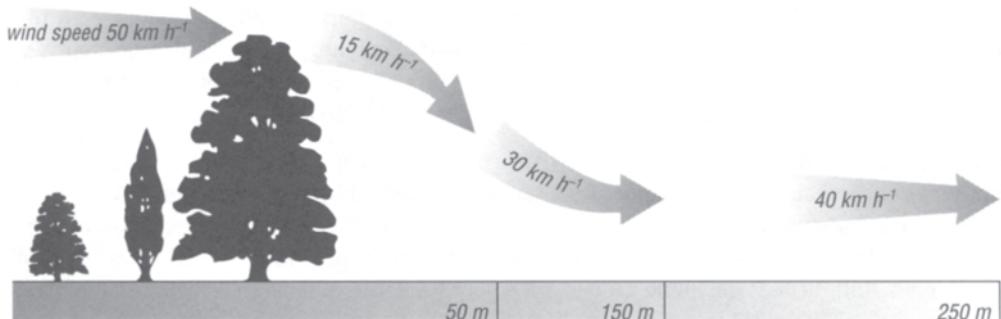


Figure 3.16 Effect of a windbreak in reducing wind speed

Erosion by water may remove more than just the soil. If the land has been tilled, sown, and fertilized, the seed and fertilizer may also be lost, possibly to the advantage of the next field downhill, but more likely to the detriment of the water body into which it drains. Wind erosion may do even worse: it may carry away young plants that have only just emerged and are not yet anchored by strong root systems. The eroded soil also causes damage. It pollutes rivers, for example, and forms sediments in reservoirs, reducing their capacity. Wind-blown soil particles can severely batter crop plants and the soil can accumulate rapidly, sometimes to considerable depth, around fences and buildings, and may cover roads.

It is somewhat scandalous that soil erosion should still be a problem of such magnitude. Ancient as farming itself, its causes and remedies are very well known. It continues where they are not applied, through ignorance, rural poverty, or indolence, or where farmers who could afford to implement remedial measures perceive the cost to them as greater than their occasional loss.

30 Mining and processing of fuels

Originally, the word ‘fossil’ described anything dug up from below ground; zoologists still describe burrowing animals as ‘fossorial’. Later, the word came to be applied to the preserved remains or traces of organisms that lived long ago (technically, more than ten thousand years ago). What we call ‘fossil fuels’ warrant their name on both grounds, but it may be more informative to describe them as ‘carbonaceous’, or ‘carbon-based’ fuels, because their combustion represents the rapid oxidation to carbon dioxide of the carbon they contain, which is an exothermic (heat-releasing) reaction.

Ordinarily, metabolic wastes and dead organisms are decomposed more or less rapidly. Most of the organisms responsible for decomposition require oxygen for respiration, however, and in anoxic environments their activities are curtailed. Under these circumstances it is possible for organic matter to become trapped, compressed beneath the weight of material that continues to accumulate above it, and subjected to rather different processes. Suitably airless environments are found, for example, in seafloor muds and below the surface of bogs and some swamps.

Plant material buried below the surface of a bog may be compressed into peat. If, later, the bog partly dries, the peat remains and can be dug for use as fuel. In some countries, such as Ireland, it is used in electricity power stations. Peat is the first stage in the formation of coal, into which it is converted by being subjected to much greater pressure and then heated: a 1 m seam of coal probably began as a 12 m layer of peat. The conditions necessary for the formation of coal occur only in the swamps found beside tropical rivers and seashores. Some of the coal being mined now formed around 400

million years ago, during the Silurian Period, but most dates from the Carboniferous, about 300 million years ago. Tectonic movements have since transported it to most parts of the world from Pangaea, the former supercontinent in which all the present continents were united and where it formed (ALLABY, 1993, pp. 143–151).

Coal and peat contain ‘volatiles’, substances that are given off as gases when the material is heated in the absence of air, and the quality of the fuel is determined by the proportion of volatiles it contains: the lower the proportion the more energy the fuel will release when burned. Peat contains more than 50 per cent volatiles, lignite (a soft, brown coal) about 45 per cent, and anthracite about 10 per cent. Anthracite is the highest quality, and very hard. Bituminous coal, the most abundant type and the one most widely used domestically, has 18–35 per cent volatiles.

Petroleum forms by a somewhat similar process. Organic material is buried by sediment, usually in a river delta, and is then trapped between two layers of impermeable rock. Many oil deposits are found beneath anticlines: rock strata that have been folded upwards into domed shapes. A similar structure occurs where a large mass of salt, deep below the surface, rises slowly through the less dense material surrounding it and the dense rock sinks to replace it. The process is called ‘diapirism’ and the salt dome it produces a ‘diapir’. Oil is often found in ‘salt-dome traps’. The material is then strongly compressed and heated. The resulting fluid fills all the pore spaces within the porous rock around it (ALLABY, 1992, pp. 162–163).

Some of the carbon and hydrogen comprising the organic matter form methane (CH_4), associated with both coal and oil. In coal mines, methane can cause fires, but when associated with oil it can be extracted and used as the fuel known as ‘natural gas’ (to distinguish it from ‘town gas’, mainly carbon monoxide (CO), obtained by heating coal and formerly an important industrial and domestic fuel).

Coal occurs in seams of varying thickness and at varying depths. There are four ways in which access can be gained to the seams and the coal extracted, illustrated in Figure 3.17. Seams at great depth are approached by sinking a shaft vertically from the surface, with associated shafts for ventilation. Most traditional British mines were shaft mines of this type, working deep seams.

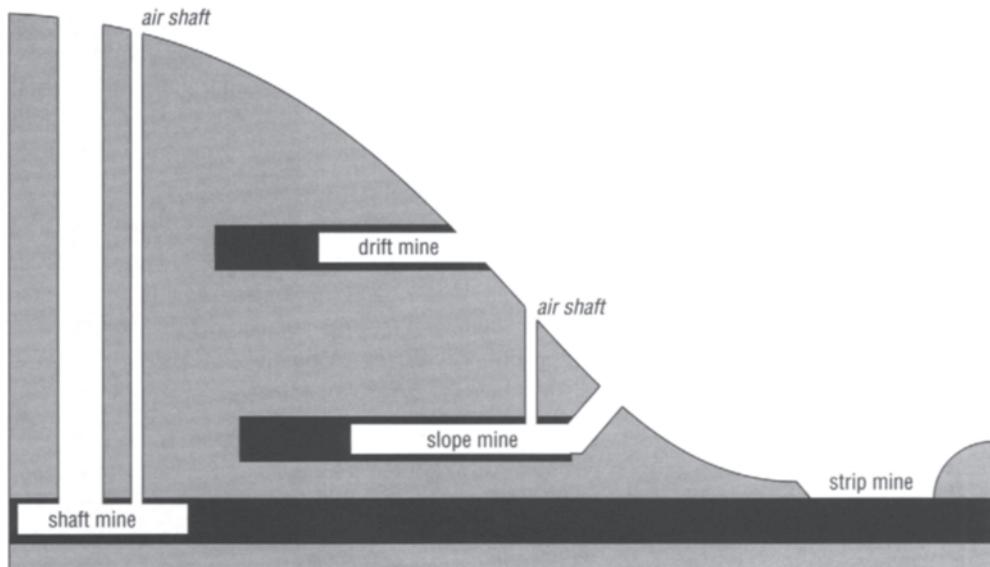


Figure 3.17 **Types of coal mines**

A slope mine approaches the seam through an angled shaft and a drift mine approaches horizontally. Where seams lie too close to the surface for a shaft to be cut to them, the overlying material is removed to expose the coal. This is an open-cast mine if it exposes a substantial part of the seam all at the same time and a strip mine if the seam is exposed and worked in sections.

All coal mining produces wastes, consisting of soil and rock that must be removed to gain access to the coal and rock mixed with the coal that must be separated from it. It is possible to store this waste, in ‘spoil heaps’, until the mine is exhausted then return it to below ground, but this is by no means the general practice and mining more commonly produces large, black spoil heaps. These are composed of finely crushed material with almost no soil and little in the way of plant nutrients. The heaps often contain large amounts of iron pyrites (FeS), producing very acid conditions ($\text{pH } 2.0\text{--}4.0$), and acid liquor, also containing metals, may leach from the heap into nearby watercourses, where it causes severe pollution. It is possible to reclaim mining spoil heaps. If they are treated with lime to reduce the acidity and soil and fertilizer supplied, a cover of grass can be established, leading in time to a more diverse plant community (MELLANBY, 1992, pp. 44–45).

Open-cast and strip mining can be even more destructive. In the past, large areas of attractive countryside were stripped of their soil (the ‘overburden’), which was dumped in large heaps, and when the seam was exhausted the site was just left, utterly devastated. In some countries this is still the practice, but in many the planning consent stipulates that when operations cease the overburden must be returned to the surface and the site restored to a state better than its original condition. The effect is not always so destructive as it may appear in the older industrial countries, such as Britain. Coal seams suitable for open-cast or strip working often occur close to deeper seams that have been mined in the past to feed nearby industries, leaving land already in a state of industrial dereliction that restoration can improve once mining has ceased. In strip mining, restoration begins long before mining ceases, the reclamation of each strip starting as soon as the extractive machinery has moved on to the next strip. Indeed, British planning regulations are now stringent, and open-cast mining has little adverse long-term effect on areas of conservation or wildlife importance.

As well as the gaseous pollutants released when it is burned, coal combustion also produces ash. This can cause disposal problems, not least because it contains heavy metals.

Oil and natural gas are held in their traps under pressure. A hole drilled through the cap rock releases the pressure and they rise to the surface. The depth of drilling determines the part of the reservoir that is tapped, since the gas lies above the oil. Figure 3.18 shows how oil and gas are held within structural traps. Because all operations are conducted from the surface and no overburden has to be removed, oil and gas mining generate no spoil heaps. Such environmental damage as they cause arises from spillages of oil around the well or in transit to the refineries where it is processed.

Coal and oil may well be forming at the present time, but at a rate much lower than that at which they are being consumed. They are, therefore, non-renewable for all practical purposes. This being so, it is commonly assumed that one day they will be economically exhausted, oil first because it is much less abundant than coal. Certainly oil is being used rapidly, and in 1994 the United States imported more petroleum and its products than it produced from its own resources, for the first time becoming a net importer (ABELSON, 1995).

Impending shortages, combined with the environmental problems arising from the combustion of fossil fuels, have stimulated a search for alternatives, but all is not necessarily as it seems. Some people suggest that valuable resources should be conserved for the benefit of future generations, but consider what has happened to the coal industry. There is probably more than 45 billion tonnes of coal lying beneath Britain and a century ago it was being mined intensively and much

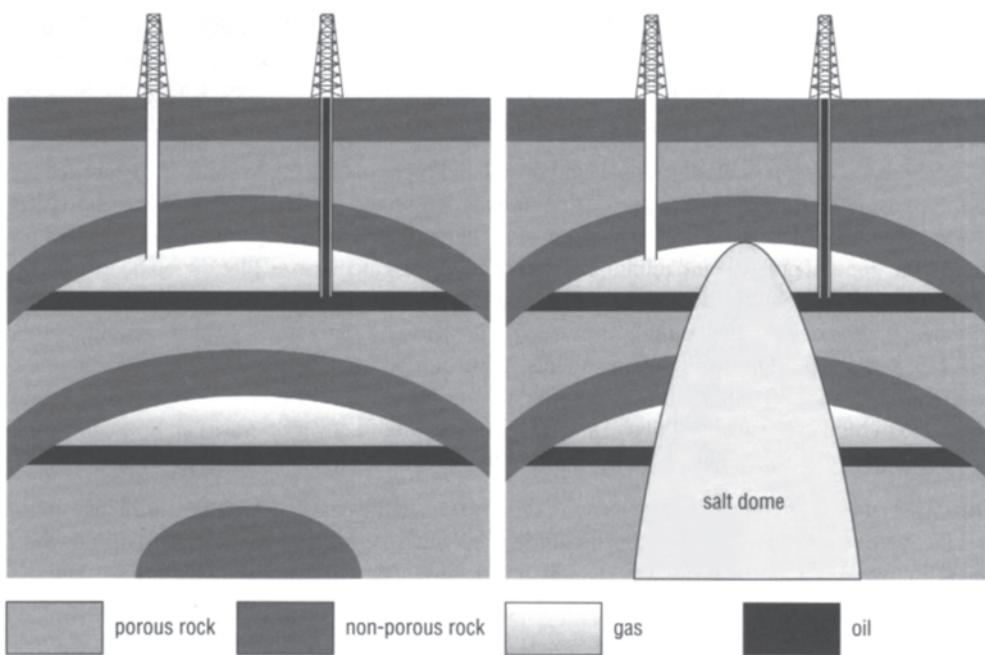


Figure 3.18 Structural oil and gas traps

of it exported. Coal is plentiful, but clearly not inexhaustible, and it might have occurred to people in the early years of this century that if the country continued to export this strategically important resource at that rate, the day would come when Britain had no choice but to import the fuel needed to power its industry and heat its homes. They might have felt it more sensible to restrict production and conserve coal for future needs. Today, matters appear rather different to those who might have benefited from such a conservation policy. In 1982, British mines produced about 125 million tonnes (Mt) of coal and the country consumed 111 Mt. In 1995, 52.6 Mt was produced and 76.2 Mt consumed (the difference was imported). Despite Britain's vast reserves, production and consumption declined, mainly because of a switch from coal to natural gas for power generation. Had the British decided years ago to restrict mining in order to conserve resources, the decision would have been wrong-headed. The loss of export earnings would have been economically damaging and the reduction in mine output would have caused unemployment. The decision would have caused real hardship, all to conserve a material for which a later generation had little use.

Relate this to the case of the United States, where coal mining and use are still increasing. There, between 1982 and 1992, coal production increased from 756 Mt to 905 Mt and consumption from 639 Mt to 808 Mt. Rising demand stimulated the search for and identification of new deposits that could be mined under the prevailing economic, political, and environmental conditions. Such identified deposits constitute 'reserves' and they increased, from 223 Mt in 1982 to 240 Mt in 1992.¹ Remembering the British experience, how sensible would it be for the United States to restrict coal mining in order to conserve coal for future generations? If the case for conserving coal is dubious, how many other materials and fuels might be of similarly doubtful future value?

Tonne for tonne, natural gas yields more energy than coal and releases fewer gaseous and no particulate pollutants. Fears of its imminent exhaustion may be unfounded. Conventional reserves may be much larger than was once supposed, and there may be a novel source: natural gas hydrate. This is a solid substance comprising a cage of water ice containing methane, technically a water clathrate of methane.

It has been found in continental permafrost regions and in the oceans at all latitudes. Methane has been recovered from several sites. It is estimated that the total amount of natural gas hydrate may be more than double that of all other fossil fuels combined (KVENVOLDEN, 1994).

Natural gas is burned in stationary installations in power stations, factories, and homes, its principal rivals for power generation being nuclear power, coal, and hydroelectric power. Solar heat and light make a minor contribution, as does wind power, the difficulty with such 'renewables' arising from their extremely dispersed nature. It would require more than 3000 of the present generation of 450 kW wind generators, for example, occupying not less than 6000 ha, to produce as much power as one modern, 1.5 GW, conventional power station, and when conditions were calm and during storms the wind generators would not function at all. In Europe, fast-growing trees, such as willows, are being grown experimentally for fuel. After harvest, the material is dried and chopped for use in power generation.

Vehicles require a liquid fuel, although they can use gas. Biomass fuels, derived from crops grown for the purpose, have the advantage of making no contribution to atmospheric carbon dioxide, because the carbon dioxide their combustion releases precisely equals the amount absorbed by the crop plants during their growth. Ethyl alcohol has been used in this way, most notably in Brazil and the United States, although it is more costly to produce than petrol. Oilseed crops are now being developed from which 'biodiesel' fuel can be obtained, one of the most promising being rape, with seeds that are 40 per cent oil. Again, production costs are high, but may be reduced through a combination of economy of scale as output increases, and genetic engineering to increase the oil content.

Electrically powered vehicles have also been the subject of much research, but they, too, present difficulties, mainly arising from the speed and range limitations imposed by the weight and power of their batteries. Nor may they be so clean as many people suppose. It has been calculated that an electric car may release into the air 60 times more lead than a car running on leaded petrol (LAVE *ET AL.*, 1995).

Fuel cells are also under development. These are devices comprising two electrodes separated by an electrolyte, which is a substance that permits the passage of ions (charged atoms or molecules) but not of electrons. A fuel containing hydrogen flows to the anode (positive electrode), where electrons are stripped from the hydrogen atoms. This leaves positively-charged hydrogen ions that diffuse through the electrolyte, while the electrons travel through an external circuit as an electric current. As the hydrogen ions arrive at the cathode (negative electrode) they are rejoined by the electrons and combine with oxygen to produce water, which is the only exhaust product. Unfortunately, fuel cells are extremely costly. The British physicist William R. Grove discovered their underlying principle in 1839 and NASA uses them to power spacecraft, but it will be several more years before they drive production cars (APPLEBY, 1999).

Energy conservation is often proposed as a partial alternative to increased exploitation of reserves or the search for new sources. If appliances and vehicles used energy more efficiently, it is argued, our demand for fuel would be correspondingly reduced. Unfortunately, the equation may not be so simple. If energy is used more efficiently, effectively it will be made cheaper and this could encourage an increase in the use of appliances to restore the balance. People would be able to obtain more use for the same price and energy consumption would not decrease. When US cars became more economical in their fuel use, during the 1970s and 1980s, consumption remained fairly constant; people drove their cars more for the same cost (INHABER AND SAUNDERS, 1994).

Nuclear power provides about 27 per cent of the electricity used in Britain and in some countries it provides much more. In Belgium, for example, it supplies 55 per cent of all electricity, in France 30 per cent, and in Lithuania 85 per cent.

Steam-driven turbines generate the electrical power and nuclear reactors produce the heat to turn water into steam. The core of the reactor comprises a structure with vertical holes or channels, some containing rods of fuel, others containing rods of cadmium or boron, all the rods being embedded in a substance called a moderator.

The fuel consists of an isotope of uranium, uranium-235 (often written as ^{235}U), that occurs as one part in 140 in natural uranium. When a slow-moving neutron collides and merges with the nucleus of an atom of ^{235}U that nucleus splits into two with the release of two or three neutrons. This is fission. If these neutrons also strike ^{235}U nuclei, the number of neutrons and nuclear fissions proliferate exponentially. This is a chain reaction and as the particles come to rest much of their energy is converted into heat.

At least one neutron from each fission must merge with a ^{235}U nucleus in order to sustain a chain reaction. Only slow-moving neutrons are able to cause fission; more energetic neutrons are not absorbed. Fission releases neutrons at many speeds and so fast-moving ones must be slowed. This is the purpose of the moderator. Different reactor designs use different moderator materials. Graphite, deuterium oxide (heavy water), and ordinary (light) water are widely used, light water being the most popular of all.

Cadmium and boron absorb neutrons, removing them from the reaction process. This means rods made from these elements can be used to regulate the speed of the chain reaction. The rods can be raised or lowered, accelerating or slowing the power output.

Surrounding the core, a coolant carries away the heat. In the most popular reactor design, the pressurized water reactor (PWR), the coolant is water under pressure. There are also designs that use boiling water. Magnox reactors, of which eight were built in Britain, the first in 1956 at Calder Hall (now Sellafield), Cumbria, use carbon dioxide as a coolant. The name ‘Magnox’ refers to the magnesium oxide alloy that is used to clad the uranium fuel rods. The advanced gas-cooled reactor also uses carbon dioxide as a coolant. Molten sodium can be used as a coolant.

Heat from the coolant passes to the water that is turned into steam to drive the turbines. In the event of a failure in the cooling system, not only would the turbines cease to function, the core would overheat, perhaps dangerously. Backup cooling systems are fitted that come into operation should this occur. These are just one of the many safety systems.

Nuclear fuel cycle

Uranium fuel for nuclear reactors is obtained from ores, in the same way as other metals. The ore is mined and milled to extract the uranium from it.

Natural uranium contains 140 parts of ^{238}U to every part of ^{235}U . It is only ^{235}U that can sustain the chain reaction necessary to generate heat. Consequently, for most reactors (but not all) it is necessary to increase the proportion of ^{235}U to about 3 per cent. This is called ‘enrichment’.

Enriched uranium is then made into fuel pellets. These are placed in metal canisters. They are then fuel rods and ready to be used.

When the ^{235}U fuel is depleted, the rods must be removed from the reactor and replaced. They are very hot and highly radioactive. They are stored under water for a time, while their temperature and level of radioactivity fall.

They are then sent for reprocessing. This operation separates fuel that can be used from the waste, including radioactive products of uranium fission. Fuel is returned for use.

The remaining waste is converted into cylindrical blocks of a glass-like solid, packaged, and sent for disposal. This involves secure storage for a number of years before final disposal. No decision has yet been made on the method to be used for final disposal, but it is most likely to be below ground in a facility constructed in a geologically stable environment.

This sequence of operations, from mining to final disposal, constitute the nuclear fuel cycle.

Chernobyl

In the course of the history of civil nuclear power there has been only one accident in which a significant amount of radioactive material was released into the environment outside the plant. That accident occurred at about 1.23 a.m. Moscow time on April 26, 1986, in the No. 4 unit of the Chernobyl nuclear power station. The station is located in the eastern part of the Belorussian-Ukrainian Woodland about 4 km from the town of Pripyat, where most of the workers from the station lived, and about 18 km from the town of Chernobyl. It was reactor of the RMBK-1000 type, using slightly enriched uranium dioxide fuel encased in cans of zirconium alloy, water as a coolant, and graphite as a moderator. The coolant water flowed through the channels into which the fuel rods were inserted. Chernobyl-4 entered service at the end of 1983.

Chernobyl-4 was to have been shut down for routine maintenance on April 25. As this happened the management planned to carry out experiments to test the safety equipment. One experiment aimed to determine how long a single turbogenerator would continue to supply sufficient power for the plant to continue operating safely once its supply of steam had been cut off and it was spinning by inertia. This necessitated switching off the reactor emergency cooling system, a procedure strictly forbidden by the authorities, to prevent it from automatically restoring the steam supply. The experiment was begun by the day shift, which included the station managers and specialists, but conducted mainly by the night shift of less qualified workers.

By about 1.20 a.m. the reactor was running at about 6 per cent of its normal power and to maintain a sufficient flow of neutrons to sustain the reaction almost all of the 211 control rods were withdrawn completely. This also broke a rule on reactor safety. The steam supply was then withdrawn from the turbine. The head of the shift realized the situation was dangerous and ordered the control rods to be reinserted. They moved very slowly, however. The number of neutrons increased suddenly (a condition called 'prompt criticality') and within one second the power surged to several hundred times its normal level. This caused the first explosion in which some of the fuel cans burst.

Coolant water then came into contact with hot fuel, producing steam that reacted with the graphite moderator. This caused the second, very much bigger explosion that blew away the top of the reactor, allowing radioactive material to escape.

More than 30 fires began. The station fire brigade arrived within five minutes and the brigade from Chernobyl a few minutes later. Helicopters were used to drop boron through the hole in the roof to capture neutrons. When the fires had been extinguished, the reactor was sealed inside a concrete casing, called a ‘sarcophagus’. Some years later, fears were expressed about the condition of the sarcophagus.

A total of 31 workers were killed by burns, falling masonry, and radiation, and about 150 people on the site suffered from radiation sickness. The radioactive plume was detected in Sweden.

Within 24 hours 47000 local residents had been evacuated and by May 7 everyone living within a radius of 30 km had been removed. In all, 116000 people from 186 settlements were evacuated. In subsequent years still more people were moved out of the affected area. Although radiation was detected further afield, outside the 30 km zone its effect on human health was predicted to be too small to be statistically detectable.

In the following years the health of exposed people was monitored closely under the leadership of the World Health Organization. Soon after the event, the most serious problem was caused by the psychological and social trauma of the accident and its aftermath of compulsory evacuation. Later there was an increase in thyroid cancer among children in Belarus.

Our production and use of fuels raises many environmental and economic questions. Posing those questions is not difficult, but finding answers to them is.

31 Mining and processing of minerals

A ‘mineral’ is a naturally occurring inorganic substance with a crystalline structure and characteristic chemical composition. Rocks are composed of minerals.

Whole rock is obtained by quarrying. Blocks of suitable rock are used for construction, in the case of slate after being split into thin sheets for roofing and cladding. Sand and gravel is also used for building, mainly of roads. Clay, won by a type of open-cast mining, is used for brick-making. High-pressure hoses (called ‘monitors’) are used to wash kaolin, or china clay, from the granite matrix in which it occurs, and it is removed as a slurry for purifying and drying. It was used originally to make fine ceramics (porcelain) but its principal use now is as a filler and whitener in paper and other materials.

Rock and building stone are quarried on a huge scale. Each year, over the world as a whole, rivers deliver to the sea about 24 billion tonnes of naturally weathered rock. Humans remove about 3 billion tonnes a year (ALLABY, 1993a, p. 150). This means we are now quarrying amounts comparable to those removed by natural processes.

Most modern quarries and open-cast mines are very large and, because they exist to detach and remove rock, cannot avoid devastating their sites. Nowadays planning consents require such sites to be restored when operations cease, but many older, abandoned quarries remain. The disfigurement they cause is not permanent, although it is only fair to point out that most older quarries were much smaller than modern ones and produced building stone, sand, or gravel in modest amounts for local use. Quarries scar the land, but they do not poison it and in time plants colonize the bare ground. An unrestored quarry site is rarely of any agricultural use and so usually it remains undisturbed and eventually may mature into a place of considerable interest to naturalists and conservationists.

Mineral mining, as opposed to the quarrying of rock, is much more disruptive, because it involves separating the desired minerals from the valueless minerals with which they are associated. The minerals themselves may be gemstones. Sapphires, oriental emeralds, and rubies are all aluminium oxides (Al_2O_3), differing from one another because of colours imparted by impurities; beryl is a compound of beryllium, aluminium, silicon, and oxygen ($\text{Be}_3\text{Al}_2\text{Si}_6\text{O}_{18}$); and diamond is a form of pure carbon. All are minerals and their high monetary value indicates their rarity: if they were common they would be cheap. When a rare substance is separated from the commoner substance containing it, a residue remains and this can cause environmental difficulties.

Metals are separated from their ores, an ore being a body of rock containing that metal in a compound, called an ore mineral, in a concentration high enough to be extracted economically. The concentration of the metal within its ore mineral may be quite high. Chalcocite (Cu_2S), for example, is 80 per cent copper and the best-quality uranite, or pitchblende (UO_2), is 85 per cent uranium, but the concentration of the metal within the ore (the rock containing the ore mineral) is very different. Iron ores are widespread and abundant, and iron is seldom extracted from ores containing less than 25 per cent of the metal, but scarcer metals commanding a high market price can be extracted economically from ores containing as little as 1 per cent or even less, for example in the case of copper. This means that in the case of iron, up to 75 per cent of the rock, and with some metals as much as 99 per cent of it, is useless waste for which some means of disposal must be found.

Mine waste cannot be returned easily to the hole whence it came. Unless the mine is exhausted, the waste would bury extractable ore, but in any case the waste no longer fits the hole. As rock, the minerals were tightly compressed; once they have been broken, crushed, and processed further to remove the desired ore mineral they consist of small particles with spaces between them. This greatly increases the bulk of the material, and there may be a truly vast amount of it. At Bingham, Utah, for example, copper is extracted from a hole 3.2 km in diameter and 900 m deep; this is large enough to accommodate two Empire State Buildings, one on top of the other, with room to spare. A lead mine in Missouri has two underground machinery repair shops 10 km apart, one with a floor area of 1.5 ha and the other 2.2 ha. Dry rock removed from these holes and expanded by processing is usually tipped to build hills of 'tailings'. Wet process residues are stored in ponds.

Tailings must be treated with respect, because the minerals they contain are in the form of tiny fragments, with a vastly increased surface area, and they are exposed to water. Subsequent chemical reactions can release very acid liquids, sometimes containing other toxic metal compounds. Dry tailings can be blown as dust, also causing contamination. Today, most governments set stringent regulations for the containment of mine tailings, so they cause little pollution, but it was not always so. In 1945, Britain had more than 1000 ha of land left totally derelict after overlying topsoil was mixed with tailings and dumped following iron-ore mining in what was otherwise good-quality agricultural land (MELLANBY, 1992, p. 47), and in south-eastern Tennessee an area of 145 km² remains barren to this day because of copper mining in the last century (RAVEN *ET AL.*, 1993, p. 331).

Mineral processing begins with separation of the ore mineral from the crushed rock. If the ore mineral is denser than the unwanted rock mixed with it, called ‘gangue’, water will separate them, the mineral being precipitated first from a suspension. Other minerals are separated by froth flotation. A compound with a strong affinity for the mineral is mixed with water and agitated to make a froth; when the crushed rock is added, the desired mineral adheres to the bubbles, the gangue sinks, and the froth is skimmed from the surface. The separated mineral is removed and dried, ready for the next stage in its preparation, leaving behind the wet gangue.

Most ores are then heated to a temperature at which the metal melts, and the elements with which it was combined react to form compounds which float above the molten metal and can be removed as ‘slag’. This is smelting and it usually proceeds as a series of chemical reactions. In the smelting of iron ore in a blast furnace, for example, the oxide ore is mixed with coke to supply carbon and limestone as a ‘flux’ that reacts to bind the slag (see Figure 3.19). The carbon is partly oxidized to carbon monoxide (CO), then oxidized further by reducing the iron oxide ($\text{Fe}_2\text{O}_3 + 3\text{CO} \rightarrow 2\text{Fe} + 3\text{CO}_2$). The impure iron may then be mixed with any of a range of alloying metals and heated again in a ‘converter’ to make steel.

Some metals are purified by electrolysis. Copper, for example, is obtained by passing an electric current through a copper sulphate solution. The anode (positive electrode) is made of ore, the cathode (negative) of pure copper. Copper ions move from the solution to the cathode and sulphate ions recombine with copper at the anode. Aluminium is also purified by electrolysis because, although it occurs as oxide ores, its affinity for oxygen is so great that heating cannot reduce it without also reducing all its impurities.

At every stage, from cutting the ore from the ground to extracting the metal from its ore mineral, the pollution risk is obvious and high. The pollution is contained by sealing tailings so dust cannot blow from them, or noxious liquors leach from them, removing gases and dust from smelters before they reach the outside air, and by treating liquid effluents.

A completely different technology now exists for extracting some metals. Bacteria that either possess the ability to isolate particular metals from their compounds or can be genetically modified to make them do so allow metals to be obtained with much less environmental disruption. *Thiobacillus ferroxidans*, for example, separates copper into an acid solution containing about 50 parts of copper per million. Sulphuric acid containing the bacteria is sprayed on to the ore rock, the liquor is collected, and the metal is removed, in this case at about one-third the cost of conventional processing, and yielding nickel as a by-product. Uranium can also be ‘mined’ in this way.

The limits to growth

In 1968 a group of 30 industrialists, economists, scientists, civil servants, and others met in Rome at the invitation of Dr Aurelio Peccei, an industrialist, to discuss ‘the predicament of man’. The meeting resulted in the formation of the Club of Rome, which eventually grew to a body with about 70 members drawn from 25 countries. Members met at intervals and collaborated to sponsor studies. Their first, launched in 1970, was the construction of a computer model to trace the consequences of interactions among five factors: population growth; agricultural production; depletion of natural resources; industrial production; and environmental pollution. The model was developed on computers at the Massachusetts Institute of Technology (MIT) by a team of 16 people led by Professor Dennis L.Meadows. Its results were published in 1972 as a non-technical account, called *The Limits to Growth*. The technical information experts needed to evaluate the methods used was published later.

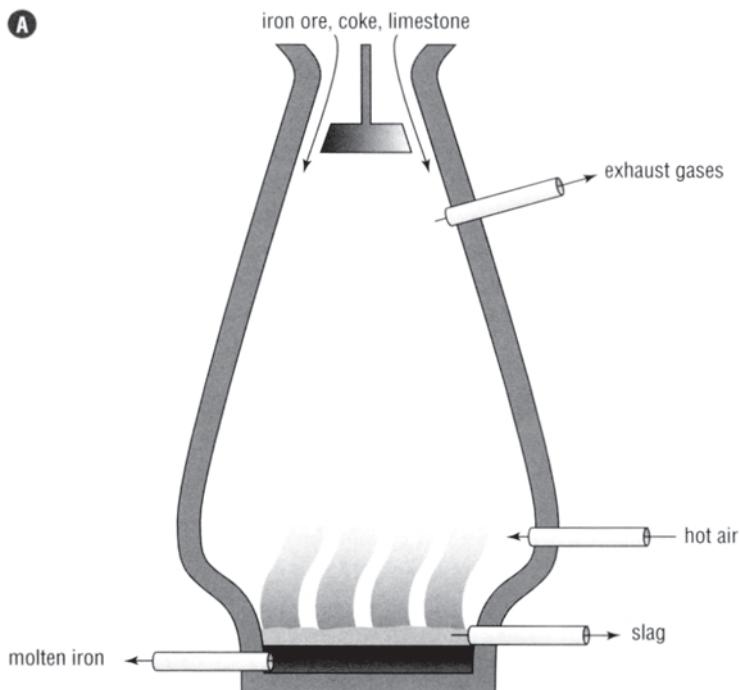
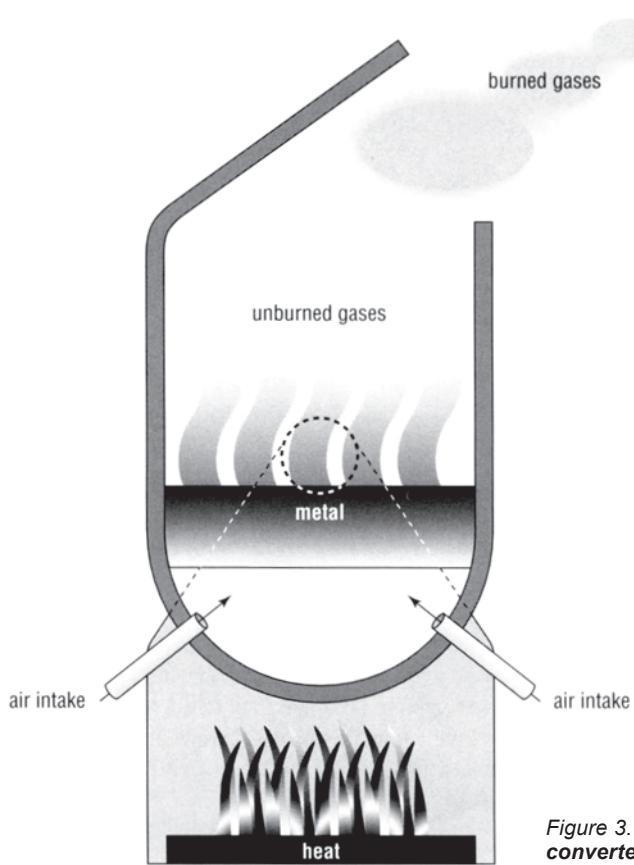
A**B**

Figure 3.19 A, Blast furnace. B, Steel converter

Pollution is only one concern arising from our use of metals; the other centres on fears of their depletion. Such fears were being expressed in the 1960s, but were stated more forcefully in the 1972 study *The Limits to Growth* (MEADOWS ET AL., 1972), which predicted the imminent exhaustion of most of the mineral resources on which we depend. The fear was based on a misunderstanding of the way the amount of a reserve is determined. The concept is economic, not physical. Figures for the reserves of any particular mineral are prepared by the mining industry primarily for its own use. They refer to the amount of the substance that has been identified and that can be profitably extracted under present conditions. They say nothing whatever about the total amount the world possesses, and changing circumstances may bring about an increase or decrease in the amount of reserves. In an apparent paradox, should consumption increase reserves may also increase to meet the demand; it was for this reason that between 1950 and 1970 global reserves of bauxite (aluminium ore) increased by 279 per cent, copper by 179 per cent, chromite (chromium ore) by 675 per cent, and tin by 10 per cent (ALLABY, 1995, pp 176–178). Historically, mineral consumption has always increased faster than the rate of population growth. World population doubled between 1950 and 1990, for example, but during the same period production of aluminium, copper, lead, nickel, tin, and zinc increased eightfold. It is now estimated that when allowance is made for anticipated population growth and economic development, mineral reserves are adequate for the next century and the environmental problems associated with their exploitation can be contained (HODGES, 1995).

Nor did the fears of exhaustion take sufficient account of the extent to which technological change renders resources obsolete. Ceramics, made from clays and sand, are now starting to replace metals for a number of uses that require tolerance of very high temperatures. Glass fibres, essentially made from silica (i.e. sand) are replacing copper cables. Orbiting satellites handle communications that were formerly transmitted by submarine cable. Electronic devices now form the basis of industrial switchgear, formerly based on mercury. Such changes, and there are many, are introduced not because of actual or anticipated shortages of the original material, but because they are superior.

The environmental implications of mining and mineral processing are well known. They can be minimized, although the restoration of mined land is sometimes difficult. In years to come, however, environmental pressures may ease. Technological advances now promise to reduce our dependence on some metals by substituting superior materials that are extremely plentiful and can be processed with much less risk of environmental harm, and by developing new, cheaper, and less disruptive extractive methods.

End of chapter summary

Without water we would die and so we may think of fresh water as the most fundamental of all the resources on which we depend. Water moves between the ocean and land, but we can manipulate the hydrologic cycle where the amount of fresh water is insufficient for our needs.

We also need soil. It is important to understand how soil forms and that it goes through a life cycle, much like a living organism. There are young, mature, old, and even senile soils. The age of a soil is directly relevant to the use that can be made of it. Fears over tropical deforestation, for example, arise partly from the realization that many tropical soils are very ancient and that this makes them inherently infertile, so alternative forms of land use may fail in the long term. Soil can be lost through erosion. The mechanisms involved in this process are well known, as are the management techniques by which they can be avoided.

Water and soil to grow food and fibre are resources essential for all animals. In addition to these, humans need industrial resources. These comprise so-called fossil fuels, rocks for building, and minerals from which metals and a range of chemical compounds are obtained. Obtaining these materials from the ground and their subsequent processing create environmental problems that must be addressed. It is usually better to prevent them by advance planning than to remedy them later.

End of chapter points for discussion

- Can shortages of water be remedied by desalination?
- What are the differences between young and old soils?
- What are the environmental effects of mineral extraction?
- What are the principal causes of soil erosion?

See also

- Weathering (section 8)
- Albedo and heat capacity (section 12)
- Greenhouse effect (section 13)
- Atmospheric circulation (section 15)
- Fresh water (section 22)
- Eutrophication (section 23)
- Salt water, brackish water, desalination (section 24)
- Irrigation, waterlogging, salinization (section 25)
- Transport by water and wind (section 27)
- Erosion (section 29)
- Farming (section 53)
- Genetic engineering (section 55)

Further reading

Biology of Freshwater Pollution, 2nd edn. C.F.Mason. 1991. Longman Scientific and Technical, Harlow, Essex. A detailed, somewhat technical explanation of the effects of discharges into rivers and lakes.

Elements: Earth. Michael Allaby. 1993. Facts on File, New York, 1993. A broad, general account of rocks and soils and the uses we make of them.

Fundamentals of Soil Science. H.D.Foth and L.M.Turk. 1972. John Wiley & Sons, New York. A fairly simple textbook on soil science.

The Human Impact on the Natural Environment. Andrew Goudie. 1986. Basil Blackwell, Oxford. Provides a clear overview of the environmental effects of industrial activity.

Only One World. Gerard Piel. 1992. W.H.Freeman, New York. A general account of the human effect on the environment, lucidly written by the former publisher of *Scientific American*.

Soil Conservation. Norman Hudson. 1971. B.T.Batsford, London. A detailed account of the mechanisms of soil erosion and techniques for preventing it.

Soils: An Introduction to Soils and Plant Growth. Roy L.Donahue, Raymond W.Miller, and John C.Shickluna. 1958. Prentice-Hall, Englewood Cliffs, NJ. Simply and clearly written, mainly for students of agriculture.

Waste and Pollution: The Problem for Britain. Kenneth Mellanby. 1992. HarperCollins, London, 1992. An assessment of the environmental effects of industrial and domestic waste disposal by possibly the leading British authority on environmental pollution.

Note

- 1 Figures from *Britannica Book of the Year 1985, 1995, and 1999*. Encyclopaedia Britannica, Chicago.

References

- Abelson, Philip H. 1995. 'Renewable liquid fuels', *Science*, 268, 995.
- Allaby, Michael. 1987. 'Environment' entry in *Britannica Book of the Year 1987*. Encyclopaedia Britannica, Chicago.
- . 1993. *Elements: Fire*. Facts on File, Inc., New York.

- . 1993a. *Elements: Earth*. Facts on File, New York.
- . 1995. *Facing the Future*. Bloomsbury Publishing, London.
- . 1998. *Dangerous Weather: Droughts*. Facts on File, Inc., New York.
- Appleby, A.John. 1999. 'The Electrochemical Engine for Vehicles', in *Scientific American*, July 1999, pp. 58–63.
- Cruickshank, James G. 1972. *Soil Geography*. David and Charles, Newton Abbot.
- Dajoz, R. 1975. *Introduction to Ecology*, 3rd edn. Hodder and Stoughton, London.
- Donahue, Roy L., Miller, Raymond W., and Shickluna, John C. 1958. *Soils: An introduction to soils and plant growth*. Prentice-Hall, Eaglewood Cliffs, N.J.
- Foth, H.D. and Turk, L.M. 1972. *Fundamentals of Soil Science*. John Wiley and Sons, New York.
- Goudie, Andrew. 1986. *The Human Impact on the Natural Environment*. Basil Blackwell, Oxford.
- Harvey, J.G. 1976. *Atmosphere and Ocean: Our fluid environments*. Artemis Press, Horsham, Sussex.
- Hodges, Carroll Ann. 1995. 'Mineral resources, environmental issues, and land use', *Science*, 268, 1305–1312.
- Hudson, Norman. 1971. *Soil Conservation*. B.T.Batsford, London.
- Hunt, Charles B. 1972. *Geology of Soils*. W.H.Freeman and Co., San Francisco.
- Inhaber, Herbert and Saunders, Harry. 1994. 'Road to nowhere,' *The Sciences*, Nov./Dec. 1994. NYAS, New York, pp. 20–25.
- Kupchella, Charles E. and Hyland, Margaret C. 1986. *Environmental Science*. Allyn and Bacon, Needham Heights, Mass.
- Kvenvolden, Keith A. 1994. 'Natural gas hydrate occurrence and issues', in Sloan, E.Dendy Jr., Happel, John, and Hnatow, Miguel A. (eds) *Natural Gas Hydrates*, Annals of the NY Acad. of Sciences vol. 715. NYAS, New York, pp 232–246.
- Lave, Lester B., Hendrickson, Chris T., and McMichael, Francis Clay. 1995. 'Environmental implications of electric cars,' *Science*, 268, 993–995.
- L'hirondel, Jean-Louis. 1999. 'Are dietary nitrates a threat to human health?', in Morris, Julian and Bate, Roger (eds) *Fearing Food*, Butterworth Heinemann, Oxford, pp. 38–46.
- Mason, C.F. 1991. *Biology of Freshwater Pollution*, 2nd edn. Longman Scientific and Technical, Harlow, Essex.
- Meadows, Donella H., Meadows, Dennis L., Randers, Jorgen, and Behrens, William W. III. 1972. *The Limits to Growth*. Earth Island, London.
- Mellanby, Kenneth. 1992. *Waste and Pollution: The problem for Britain*. HarperCollins, London.
- Ministry of Agriculture, Fisheries and Food and Department of Agriculture and Fisheries for Scotland. 1968. *A Century of Agricultural Statistics; Great Britain 1866–1966*. HMSO, London.
- Moore, Peter D. 1995. 'Too much of a good thing', *Nature*, 374, 117–118.
- Piel, Gerard. 1992. *Only One World*. W.H.Freeman, New York.
- Raven, Peter H., Berg, Linda R., and Johnson, George B. 1993. *Environment*. Saunders College Publishing, Orlando, Florida.
- Reinhold, Robert. 1992. 'The lingering US drought,' *Britannica Book of the Year* 1992. Encyclopaedia Britannica, Chicago, p. 168.
- Royal Commission on Environmental Pollution. 1979. *Seventh Report: Agriculture and Pollution*. HMSO, London.
- Smith, George E. 1967. 'Fertilizer nutrients as contaminants in water supplies,' in Brady, Nyle C. (ed.) *Agriculture and the Quality of Our Environment*, AAAS, Washington, DC. p. 177.
- Tolba, Mostafa K. and El-Kholi, Osama A. 1992. *The World Environment 1972–1992*. Chapman and Hall, London, on behalf of UNEP.
- Winter, E.J. 1974. *Water, Soil and the Plant*. Macmillan Press, London.