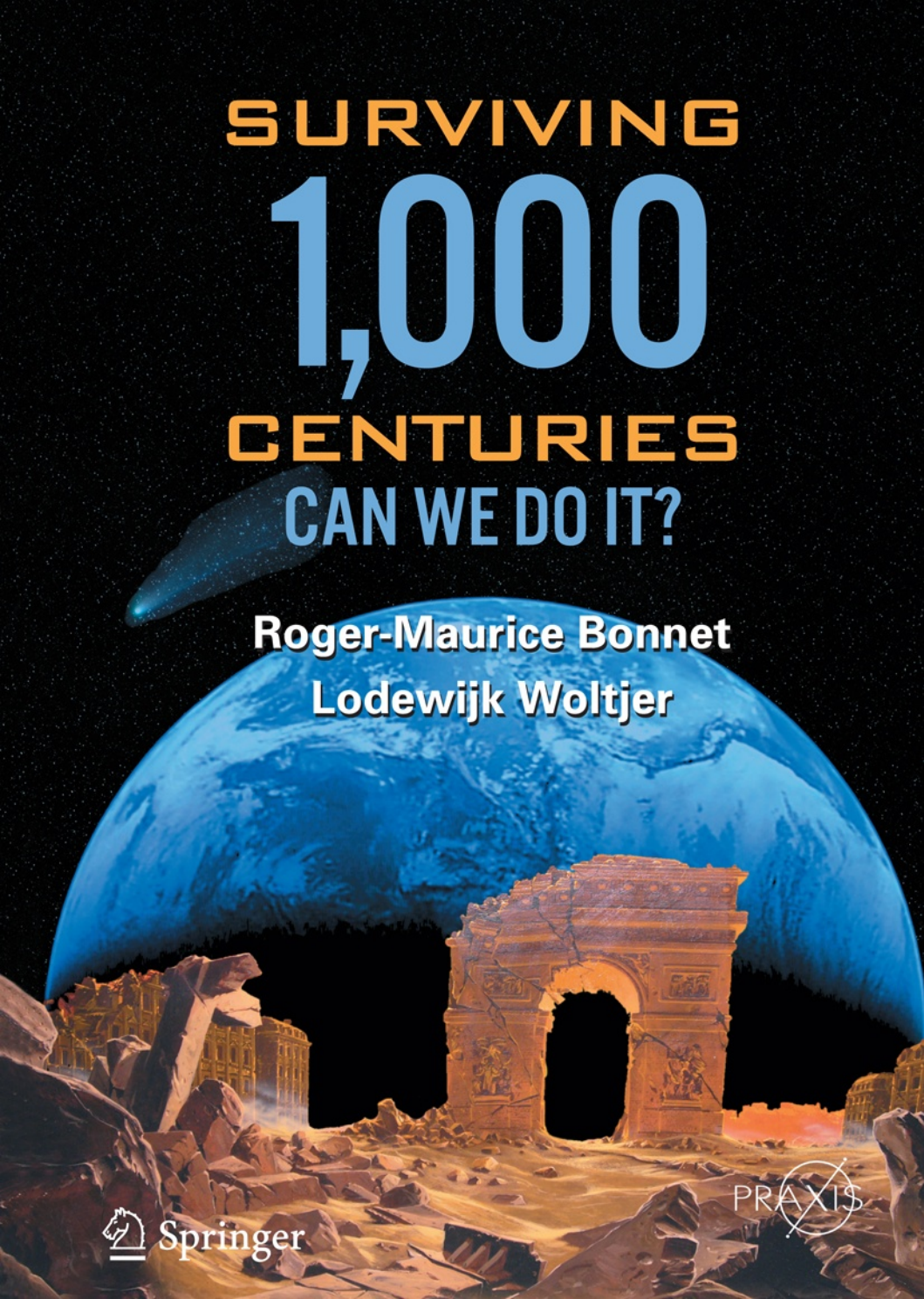
SURVIVING 1,000 CENTURIES

CAN WE DO IT?

Roger-Maurice Bonnet
Lodewijk Woltjer



Springer
PRAXIS

Surviving 1,000 Centuries

Can we do it?

Roger-Maurice Bonnet and Lodewijk Woltjer

Surviving 1,000 Centuries

Can we do it?

Springer

Published in association with
Praxis Publishing
Chichester, UK

PRAXIS

Dr Roger-Maurice Bonnet
President of Cospar
Executive Director, The International
Space Science Institute (ISSI)
Bern
Switzerland

Dr Lodewijk Woltjer
Observatoire de Haute Provence
Saint-Michel l'Observatoire
France

Credit for the cover photo-montage:
Arc de Triomphe painting credit: Manchu/Ciel et Espace.
Earth crescent: first high-definition image of the Earth obtained on board the KAGUYA lunar explorer (SELENE) from a distance of about 110,000 km away. Credit: Japan Aerospace Exploration Agency (JAXA) and NHK (Japan Broadcasting Corporation).

SPRINGER–PRAXIS BOOKS IN POPULAR SCIENCE
SUBJECT *ADVISORY EDITOR*: Stephen Webb B.Sc., Ph.D.

ISBN 978-0-387-74633-3 Springer Berlin Heidelberg New York

Springer is a part of Springer Science + Business Media (*springer.com*)

Library of Congress Control Number: 2008923444

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

© Copyright, 2008 Praxis Publishing Ltd.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: Jim Wilkie
Editor: Alex Whyte
Typesetting: BookEns Ltd, Royston, Herts., UK

Printed in Germany on acid-free paper

Contents

List of Illustrations xi
Foreword xv
Preface xvii
Acknowledgments xix

1 Introduction 1
1.1 Why a hundred thousand years? 1
1.2 People and resources 5
1.3 Management and cooperation 7
1.4 The overall plan of the book 9
1.5 Notes and references 11

2 A Brief History of the Earth 13
2.1 The age of the Earth 13
2.2 Geological timescales 16
2.3 The formation of the Moon and the Late Heavy Bombardment 18
2.4 Continents and plate tectonics 23
2.4.1 Continents 23
2.4.2 Plate tectonics 24
2.4.3 The Earth's magnetic field 28
2.5 Evolution of the Earth's atmosphere 31
2.6 Life and evolution 35
2.6.1 The early fossils in the Archean 35
2.6.2 The Proterozoic and the apparition of oxygen 37
2.6.3 The Neo-Proterozoic: the Ediacarans and the 'snowball earth' 38
2.6.4 The Phanerozoic, life extinctions 43
2.7 Conclusion 48
2.8 Notes and references 48

3 Cosmic Menaces 53
3.1 Introduction 53
3.2 Galactic hazards 54
3.2.1 The death of the Sun 57
3.2.2 Encounters with interstellar clouds and stars 57
3.2.3 Supernovae explosions, UV radiation and cosmic rays 59
3.2.4. Gamma-ray bursts and magnetars 60

3.3 Solar System hazards 62
3.3.1 Past tracks of violence 62
3.3.2 The nature of the impactors: asteroids and comets 66
3.3.3 Estimating the danger 73
3.3.4 The bombardment continues 76
3.3.5 Mitigation measures 80
3.3.6 Deviation from the dangerous path 80
3.3.7 Decision making 84
3.3.8 Space debris 85
3.4 Conclusion 89
3.5 Notes and references 89

4 Terrestrial Hazards 93
4.1 Introduction 93
4.2 Diseases 95
4.2.1 How old shall we be in 1,000 centuries? 98
4.2.2 How tall shall we be in 1,000 centuries? 100
4.3 Seismic hazards: the threat of volcanoes 102
4.3.1 Volcanoes and tectonic activity 102
4.3.2 The destructive power of volcanoes 106
4.3.3 Volcanoes and climate change 109
4.3.4 Forecasting eruptions 112
4.4 Seismic hazards: the threat of earthquakes 115
4.4.1 Measuring the power of earthquakes 119
4.4.2 Earthquake forecasting 120
4.4.3 Mitigation against earthquakes 125
4.5 Tsunamis 125
4.5.1 What are they? 125
4.5.2 The 26 December 2004 Sumatra tsunami 127
4.5.3 Forecasting tsunamis and mitigation approaches 128
4.6 Climatic hazards 132
4.6.1 Storms: cyclones, hurricanes, typhoons, etc. 132
4.6.2 Floods 137
4.6.3 Droughts 142
4.7 Conclusion 146
4.8 Notes and references 148

***5* The Changing Climate** 153
5.1 Miscellaneous evidence of climate change 153
5.2 The global climate system 156
5.3 Climates in the distant past 160
5.4 The recent ice ages 163
5.5 Recent climate 171
5.6 Changes in the Sun 174
5.7 Volcanic eruptions 176

5.8 Anthropogenic CO_2 177
5.9 Interpretation of the recent record 178
5.10 The ozone hole 179
5.11 Notes and references 182

6 Climate Futures 187
6.1 Scenarios for future climates 188
6.2 Geographic distribution of warming 194
6.3 Sea level 197
6.4 The 100,000-year climate future 201
6.5 Doubts 205
6.6 Consequences of climate change 206
6.7 Appendix 207
6.7.1 The four main SRES scenarios 207
6.8 Notes and references 209

7 The Future of Survivability: Energy and Inorganic Resources 213
7.1 Energy for 100,000 years 213
7.1.1 Energy requirements for the 100,000-year world 215
7.1.2 Minor energy sources for the long-term future 217
7.1.3 Wind energy 219
7.1.4 Solar energy 221
7.1.5 Biofuels 223
7.1.6 Nuclear energy 225
7.1.7 Fusion energy 228
7.2 Energy for the present century 232
7.2.1 Fossil carbon fuels 232
7.2.2 Electricity and renewables 236
7.2.3 From now to then 236
7.3 Elements and minerals 238
7.3.1 Abundances and formation of the elements 238
7.3.2 The composition of the Earth 241
7.3.3 Mineral resources 242
7.3.4 The present outlook 244
7.3.5 Mineral resources for 100,000 years 245
7.3.6 From now to then 250
7.4 Conclusion 250
7.5 Notes and references 250

8 The Future of Survivability: Water and Organic Resources 253
8.1 Water 253
8.1.1 The water cycle 254
8.1.2 Water use and water stress 255

8.1.3 Remedial measures 257
8.1.4 Water for 100,000 years 260
8.1.5 From now to then: water and climate change 262
8.2. Agriculture 263
8.2.1 Increasing productivity 263
8.2.2 Present and past land use 265
8.2.3 Population 266
8.2.4 Agricultural land and production 266
8.2.5 Irrigation 267
8.2.6 Fertilizers and pesticides 267
8.2.7 Top soil 267
8.2.8 Agriculture for 100,000 years 269
8.2.9 From now to then 271
8.3 Forests and wilderness 271
8.3.1 Deforestation 273
8.4 Conclusion 276
8.5 Notes and references 276

9 Leaving Earth: From Dreams to Reality? 281
9.1 Introduction 281
9.2 Where to go? 282
9.2.1 The case of Venus 284
9.2.2 The case of Mars 288
9.2.3 Other worlds 294
9.2.4 Interstellar travel 297
9.2.5 Space cities? 299
9.3 What to do with the Moon? 300
9.3.1 The Lunar Space Station 301
9.3.2 The Moon as a scientific base 303
9.3.3 The Moon for non-scientific exploitation 303
9.3.4 Resources from outside the Earth–Moon system: planets and asteroids 306
9.4 Terraforming the Earth 308
9.4.1 Absorbing or storing CO_2 308
9.4.2 Cooling down the Earth 309
9.5 Conclusion 311
9.6 Notes and references 311

10 Managing the Planet's Future: The Crucial Role of Space 315
10.1 Introduction 315
10.2 The specific needs for space observations of the Earth 316
10.2.1 The Earth's interior 316
10.2.2 Water: the hydrosphere and the cryosphere 319
10.2.3 The atmosphere 323
10.2.4 The biosphere 327

10.3 The tools and methods of space 329
10.3.1 The best orbits for Earth observation 330
10.3.2 Geodesy and altimetry satellites: measuring the shapes of the Earth 331
10.3.3 Global Positioning Systems 337
10.3.4 Synthetic Aperture Radars 339
10.3.5 Optical imaging 347
10.3.6 Remote-sensing spectroscopy 350
10.3.7 Radiometry 354
10.3.8 Monitoring astronomical and solar influences 357
10.4 Conclusion 362
10.5 Notes and references 363

11 Managing the Planet's Future: Setting-Up the Structures 367
11.1 Introduction 367
11.2 The alert phase: need for a systematic scientific approach 368
11.2.1 Forecasting the weather: the 'easy' case 368
11.2.2 The scientific alert phase: the example of the IPCC 372
11.2.3 Organizing the space tools 376
11.3 The indispensable political involvement 381
11.3.1 The crucial role of the United States, China and India 381
11.3.2 A perspective view on the political perception 384
11.3.3 The emotional perception: the scene is moving 393
11.4 Conclusion: towards world ecological governance? 397
11.5 Notes and references 399

12 Conclusion 403
12.1 Limiting population growth 403
12.2 Stabilizing global warming 405
12.3 The limits of vessel-Earth 406
12.4 The crucial role of education and science 407
12.5 New governance required 408
12.6 The difficult and urgent transition phase 410
12.7 Adapting to a static society 411
12.8 Notes and references 413

Index 415

List of Illustrations

1.1	Rise in human population	4
2.1	Geological epochs	18
2.2	Accretion rate on the Moon	22
2.3	Episodes of crustal growth	24
2.4	The Pangea super-continent	25
2.5	Structure of Earth's magnetic field	29
2.6	Evolution of Earth's magnetic field	30
2.7	The faint young Sun problem	34
2.8	Tree of life	37
2.9	Grypania	38
2.10	Ediacaran fossils	40
2.11	Earth's history	41
3.1	The heliosphere	54
3.2	The Earth's magnetosphere	55
3.3	Earth viewed from the ISS	55
3.4	Collision of two galaxies	56
3.5	NO_x production by cosmic rays	58
3.6	The Crab Nebula	59
3.7	The Moon's South Pole	63
3.8	Sample of asteroids	64
3.9	Chicxulub crater in Yucatan	65
3.10	Double impact crater	66
3.11	Oort Cloud, Kuiper Belt, Asteroid belt	67/68
3.12	Known Near-Earth Asteroids	69
3.13	Asteroid Itokawa	71
3.14	Nucleus of Halley's Comet	71
3.15	NASA's Deep Impact Mission	72
3.16	Fragmentation of Comet Shoemaker-Levy 9	73
3.17	ESA's Rosetta probe	81
3.18	Path of risk of Apophis asteroid	84
3.19	Orbital debris	86
3.20	Number of Low Earth Orbit objects	88
4.1	Mortality from catastrophes	94
4.2	Health workers and disease burden	95
4.3	Economic losses from disasters	96
4.4	Main causes of death	97
4.5	Deaths for selected causes	99

4.6	Tectonic plates	103
4.7	Distribution of volcanoes	104
4.8	The interior of the Earth	105
4.9	Lake Toba	108
4.10	The Pinatubo eruption	111
4.11	Map of major earthquakes	113
4.12	Types of seismic waves	117
4.13	Activity related to Sumatra earthquake	123
4.14	Propagation of the 2004 Sumatra tsunami	129
4.15	Map of DART stations in the Pacific	130
4.16	The Katrina hurricane	133
4.17	Costs of US weather disasters	135
4.18	Deaths from tropical cyclones	135
4.19	Predicted changes in tropical cyclones	136
4.20	Mosaic of the Deluge	139
4.21	Aqua alta in Venice	140
4.22	The 2003 heat wave in Europe	145
5.1	Little Ice Age	155
5.2	Retreat Muir glacier	156
5.3	Floating ice in Antarctica	157
5.4	Thermohaline circulation	159
5.5	EPICA temperatures in Antarctica	165
5.6	Temperatures and CO_2 in Vostok core	166
5.7	Orbit of the Earth	167
5.8	Temperatures in Greenland and Antarctica	169
5.9	Global temperature from 1880	170
5.10	Distribution warming 2005–2007	172
5.11	Northern hemisphere warming AD 200–2000	173
5.12	Sunspots	175
5.13	Solar irradiance 1978–2007	175
5.14	CO_2, CH_4 and N_2O 1400–2100	178
5.15	Antarctic ozone hole 2007	181
6.1	Climate forcings 1750–2100	190
6.2	Simulated snowfall in a model	196
6.3	The northwest passage	200
6.4	Past and future insolation at 65°N	202
6.5	IPCC scenarios	208
7.1	Global distribution of windspeeds	221
7.2	Fusion energy, ITER	231
7.3	Current energy production and supply	235
7.4	Elements in the Sun	239
7.5	Elements in the Earth's crust and oceans	247
8.1	The hydrological cycle	254
8.2	River runoff and water withdrawals	255
8.3	Aral Sea	259

8.4	Distribution of land use	265
8.5	Deforestation of tropical forests	275
9.1	Venus, Earth, Mars and Titan	283
9.2	The habitable zone	284
9.3	Fluvial features on Mars	289
9.4	Water-ice in Vastitas crater	290
9.5	Obliquity and insolation on Mars	291
9.6	Europa imaged by Galileo spacecraft	295
9.7	Earth rising above lunar horizon	302
10.1	South Atlantic anomaly	318
10.2	Sea level rise 1993–2006	321
10.3	Regional distribution of sea level trends	322
10.4	Altitude variation of atmospheric temperature	323
10.5	Global ozone changes 1964–2002	325
10.6	Desertic aerosols over African coast	326
10.7	Phytoplankton bloom in Baltic Sea	328
10.8	Global biosphere	329
10.9	Geoid observed by GRACE	331
10.10	Water flow through Amazon	333
10.11	Map of ocean floor	335
10.12	Rivers and swamps in Amazon basin	336
10.13	The 30 Galileo satellites	338
10.14	SAR imaging geometry	339
10.15	Flooding in Bangladesh	341
10.16	Oil spill off Spanish coast	342
10.17	Landslide in Slovenia	344
10.18	SAR image of Mount Fuji	345
10.19	Interferometric map of Bam earthquake	345
10.20	Three-dimensional view of Etna	346
10.21	Subsidence of Venice	346
10.22	Land cover over Cardiff	348
10.23	Wildfires in California in 2007	349
10.24	Opium poppy parcels in South-East Asia	350
10.25	Global ozone forecasts	351
10.26	Clear sky UV index in 2006	352
10.27	Tropospheric column density of NO_2	353
10.28	Column density of methane	354
10.29	Earth's energy budget	355
10.30	Solar irradiance at different levels	356
10.31	Map of sea-surface temperature	356
10.32	EUV images of solar cycle	357
10.33	Integrated solar UV irradiance	358
11.1	Improvements of weather forecasts	369
11.2	Three-monthly predictions El Niño	371
11.3	Global Observing System	379

11.4 Evolution of tropospheric NO_2 columns 383
11.5 Effects of Montreal Protocol 385
11.6 Evolution of ozone 1980–2050 386

Foreword

This is a fascinating book, but nevertheless comes as a great surprise to me. The authors, two eminent physicists, are confining their text to a timescale of 100,000 years, but for astrophysicists the timescales of cosmic objects such as the stars, the galaxies, and the Universe are more commonly expressed in millions or billions of years. As scientific managers the authors have been responsible for projects with timescales of only a decade, but in this book they are considering the future of our planet within a time period of one thousand centuries, which is long compared to projects but short compared to astronomical objects. All important problems relevant for this timescale – cosmic menaces, natural hazards, climate changes, energy, and resources – are covered and very carefully analyzed.

I have known both authors for some 50 years. I met Lo Woltjer for the first time in 1959 when we were neighbors in Einstein Drive at the Institute for Advanced Studies in Princeton, New Jersey, and I first saw Roger Bonnet in 1962 in the Sahara desert, where we launched sounding rockets together to investigate the upper atmosphere.

Both authors have made important contributions to our understanding of cosmic objects and both were influential managers within the European Research Organizations, the European Southern Observatory, and the European Space Organizations.

Over recent decades there have been several attempts to predict the future of our planet, but these approaches were concerned with a timescale of only 10 or 20 years, such as *Limits to Growth* of the 'Club de Rome,' or covering a period of 50 to 100 years in scenarios for the climate.

However, the analysis in this book demonstrates that it is very important to prepare plans today for the survival of our society in the distant future. The important messages are: the increases in global population must be drastically limited, and we must make plans for our energy resources over the next 100,000 years.

I think this book gives an optimistic outlook to the future rather than the pessimistic view that is more commonly expressed today. The problem is not so much the long-time future, but the transition phase from our present state to our distant future. The authors show clearly how important the stabilization of global warming is for our survival. If the world population does not exceed 11 billion, a reasonable level of comfort would be possible for at least 100,000 years, as sufficient renewable energy should be available together with fusion energy. This, at least, is the hope of these two astrophysicists, although the adequacy and

acceptability of fusion has yet to be proven. As a consequence of the detailed analysis in this book, the efforts on fusion research should not be reduced but strengthened and increased.

I hope that this book will be read by those who have political responsibility in the various countries on the globe. As most of them feel only responsible until their next election, there remains the open question of who is willing to start the right initiative as required by the authors.

Reimar Lüst
Max-Planck-Institut für Meteorologie, Hamburg

Preface

It might look strange that we have decided to write this book about the only planet in our Solar System that our telescopes do not observe. We both have been in charge of the most important European programs in astronomy and space science in the latter part of the previous century and we might not be recognized as the most competent people to deal with the state and future of our planet and the civilization that inhabits it. Having previously dealt with the entire Universe why have we now decided to turn our eyes to that small sphere of rock, water and gas on which we live?

Since we were born, the population of the Earth has multiplied by more than a factor of 3. In the meantime, science has evolved at a pace never attained before: antibiotics were invented; the atomic and hydrogen bombs were developed; the structure of matter has been nearly totally deciphered; the dream of exploring space became true in 1957 with the launch of Sputnik 1; the largest telescopes on Earth have changed our view and perception of the Universe and of its evolution; and information technology has revolutionized the lives of all people on Earth. In the meantime, the pollution of light has forced us to explore the highest and most isolated mountains on Earth on which to install our telescopes and continue with our observations. The brightness of the sky in radio waves has also been multiplied by more than 4 orders of magnitude in the last 60 years as a result of all the emissions from radio communications and television – so much so that we are now thinking of installing our radio telescopes on the hidden face of the Moon, which offers the best protection against the dazzling radio light of modern civilization.

Even if we had not wished to worry about the Earth, the state of our planet has forced us to change the ways in which we conduct our work. As we unravel the secrets of such planets as Venus and Mars – and the numerous others that we find orbiting stars in our milky way – it is impossible not to look at our own planet and ask ourselves whether it will be capable of continuing to accommodate life and to resist the tremendous changes that humans impose on its evolution, surpassing the natural changes. One of the recurring questions that people would like astronomers to answer is whether those newly discovered distant planets also harbour life. Logically, we therefore ask ourselves how long will life, as we know it, continue to exist on the only planet that we know for certain is inhabited? In other words, can we survive; and for how long?

We have left our successors the means to exploit all that we have devoted a substantial part of our life to build; it is up to them to pursue these developments further. Having retired on the Olympus of our success, we therefore decided to

look down from our lofty peak and consider our planet, the Earth, and in 2002 we began to write this book. It was only later that we confined our exercise to within the next 1,000 centuries – an option we justify in the opening pages of the book. This is a ridiculously small lapse of time, equivalent to less than 2 seconds if the age of the Earth were set equal to 24 hours. From the time we started, the planet has constantly changed. In those six years, the average temperature has risen by a further 0.06°C, the sea level has risen by nearly 2 cm, and 45 million hectares of forest have disappeared. In the meantime our life expectancy has gained more than 18 months, increasing the proportion of elderly people in the world, clearly announcing the need to reorganize our societies. How long can these societies last? What conditions must we fulfill? What options must we choose if we are to survive for a further 100,000 centuries? We have attempted to give some answers to these questions, being aware that our analysis needs to be refined and pursued in several areas.

It is certainly easier to fix targets, but less obvious to define how to go from now to then and reach these targets. In the book, we try to outline these transitions be it for energy, mineral resources water, agriculture or land use. The issue of global warming, which became so visible through the work of the IPCC during the time of writing this book, has strongly influenced our reflection. It deserves constant monitoring and imposes already difficult political and societal choices. One condition, however, seems to stand above all. These transitions will be difficult, even traumatic but easier to go through if the need to do so is well understood. In that respect, enunciating the most precise knowledge of the status of the planet is an absolute necessity. This can only be obtained through a thorough and complete scientific evaluation, involving a complex set of measurements and observations from the ground and in orbit, and a lot of modeling and computations. But this is not enough.

It is of the utmost importance that everyone should understand the problems we are facing and accept the changes made mandatory to adapt to new ways of living. That the two of us, as scientists, advocate more science and more education is nothing exceptional. The opposite attitude would be. We are deeply convinced that the Earth can offer accommodation to its future population only if those responsible understand that its limited resources require the development of socio-economical systems that have to be implemented soon – and the sooner the better!

R.-M. Bonnet
L. Woltjer

Acknowledgments

This book would not have been written without the help and support of many people and it is our pleasant duty to thank them here and acknowledge their contributions, which have helped us to make this book better.

Lo Woltjer acknowledges the Observatoire de Haute-Provence in particular, Mira Veron and the Osservatorio Astrofisico di Arcetri, where parts of this book were written, for their hospitality. He also wishes to thank Daniel and Sonia Hofstadt in whose hospitable villa in the magnificent scenery of Lago Rupanco in Chile some of thoughts expressed in this book were developed. Thanks are also due to Claude Demierre for his kind help with the graphics.

Roger-Maurice Bonnet would like to express his warmest thanks to Bernhard Fleck, Einar Erland and Stephen Briggs at ESA, to the personnel of ISSI for their constant help and their kindness in providing all the material and intellectual environment that have made his work much easier, more pleasant and better documented. Special thanks go in particular to Silvia Wenger, Saliba Saliba, Yasmine Calisesi, Iremla Schweizer and Brigitte Fassler. A particular thank-you goes also to André Balogh, not only for his help in providing and improving some of the graphics, but also for his advice in the course of selecting the editor of the book. He acknowledges also the hospitality of the Institut d'Astrophysique de Paris where part of the book was prepared and written and in particular the contribution by Geneviève Sakarovitch, who is responsible for the library.

We are both particularly grateful to those who carefully reviewed the manuscripts of the various chapters and encouraged us in pursuing this work, in particular, Professors Johannes Geiss and Lennart Bengttson, Dr Oliver Botta at ISSI, Dr Eduardo Mendes-Pereira at the Institute National de la Recherche Agronomique and Dr Jacques Proust at Clinique de Genolier. We are indebted to all our former collaborators, colleagues and members of the scientific community who have granted us permission to use and reproduce some of their results, unpublished documents or illustrations.

We would like to express our appreciation to several organizations whose copyright policy is particularly user-friendly and has allowed us to use a superb and rich iconography, in particular, ESA, CNES, NASA, JPL, GSFC, JAXA, USGS, NOAA, WMO, WHO, and ECMWF.

We both wholeheartedly thank Ulla Demierre Woltjer for her continuing support, encouragement and assistance during all phases of the writing of this book.

And last, but not least, we would like to thank Alex Whyte for his invaluable help in the editing of the text and Clive Horwood and his team at Praxis Publishing for their expert guidance through the various stages of production.

1 Introduction

Progress has often been delayed by authors, who have refused to publish their conclusions until they could feel they had reached a pitch of certainty that was in fact unattainable.

Charles Galton Darwin, *The Next Million Years*

In this book we study the physical circumstances that will shape the long-term future of our civilization. Why should we be interested? Perhaps we may have an idle curiosity for how the future will look. But there is more. The distant future will be very much constrained by what is physically possible and what is not. This may help us to select among our present options those that are viable in the future. If we were to follow a path that is a dead end, we may first have to undo the damage before we can follow one that shows more promise for the future, assuming this still to be possible. As an example, we all know that oil and gas will become exhausted in the not too distant future and also that the burning of these will lead to serious consequences for the Earth's climate, though these may not be immediately obvious. So, in the long run we shall have to find alternative sources of energy. That being the case, is it not better to start work on these alternatives now than invest all our efforts in augmenting the oil supply only to discover later that it is not enough and that irremediable damage has been done to the environment?

1.1 Why a hundred thousand years?

What is the meaning of our 'long-term future'? Is it a century, a millennium or a million years? Insightful studies have been made of future developments during the coming decades. For example, McRae [1] in 1994 published *The World in 2020* in which he outlined the anticipated developments in the world at large and the waxing and waning of the regional powers with also brief discussions of future energy resources. On such a timescale it is possible to predict the future on the basis of present trends and the assumption that no major conflict, such as a nuclear war or other unexpected upheaval, completely changes the world. Now, at the halfway point, McRae's forecasts were on the whole remarkably to the point.

On a timescale of a hundred years, or three human generations, predictions of political developments become rapidly more uncertain. However, our under-

standing of the natural world has been rapidly improving and has allowed longer term predictions to be made about climate and natural resources, even though much quantitative uncertainty remains. As a result, various international and national organizations are now making projections of energy and resource availability to the end of the present century. Also the Intergovernmental Panel on Climate Change (IPCC) makes estimates of climate developments over the same period. The level of confidence in some of these predictions has become sufficient for governments to take them into account in developing policies.

A much longer time frame had been considered in 1953 by C.G. Darwin (a grandson of Charles Darwin) in his book entitled *The Next Million Years* [2]. As Darwin stated, forecasts for a brief period into the future are very hard to make because large fluctuations due to war or disasters may have major effects. A longer period is needed to even out such events and to analyze the nature of an equilibrium that may be reached. Darwin's book was almost entirely qualitative and sociological. Two theses were proposed: that the world would run into a Malthusian disaster if population were not stabilized, and that adequate resources of metals, etc., would only be obtainable if a large supply of energy was available. He thought that the latter could result from nuclear fusion – the same process that has kept the Sun shining for several billion years.

Modern humans developed in Africa 150,000–200,000 years ago and emerged from Africa some 40,000–60,000 years ago, rapidly populating much of the world [3]. These people were very much like us. The art they produced seems familiar, as does their physical appearance. So, in our book we shall ask the question: Can we imagine a future in which we are just at the midpoint of the development of modern humans?, or formulated differently: Are the physical circumstances on Earth such that the duration of the society of modern humans can be doubled? After all it would be regrettable if we were to discover that we were already at the end of the road of a process that had such promising beginnings, though an uncertain outcome. If we go back a million years, the situation is different. Some tools were made, but art was largely absent, and significant cognitive evolution took place thereafter. So we shall set the time frame over which we project the physical circumstances for the future society to 100,000 years, at which time the natural evolution of the human race has not had too large an influence. Of course, an acceleration of our evolution by genetic manipulation might well be a possibility, but its effects are at the moment unforeseeable. Over such a timescale, the Earth will not alter much; the continents may move about and mountain ranges may rise and fall but these events occur far too slowly to change the overall geography. However, the climate and sea level may change.

We are not the first to express concern about the long-term viability of the Earth. In fact, it has inspired legislation in several countries. In the United States the law specifies that nuclear waste has to be stored safely for 10,000 years, and in the controversies about the Yucca Mountain storage site the desirability of prolonging this period has been stressed [4]. In Sweden the most long-lived radioactive waste should be placed in a depository which is guaranteed to be

secure for 100,000 years [5]. Such laws show that there is nothing exotic about being worried about the well-being of our descendants that far downstream. Of course, the long-term future of human life on Earth depends on much more than the nuclear issue. It is just the popular fear of radioactivity that has given that particular issue its prominence.

It is generally agreed that the development of the world should be 'sustainable', but what this means quantitatively is less clear. According to the Brundtland Report [6] to the United Nations, sustainable development implies 'that it meets the needs of the present without compromising the ability of future generations to meet their own needs'. Few would disagree with this laudable aim, but what it means is far from obvious. The pessimists look at the current dwindling of finite non-renewable resources and claim that present consumption levels are unsustainable. The optimists tell us that it suffices to leave our descendants the knowledge required to live well after the exhaustion of current resources, because technological fixes will solve the problems. The optimists have a case when they say that technological developments related to solar energy may cure our energy problems. But the pessimists also have a case when they argue that the loss of nature and of biodiversity is irreversible. Therefore, a deeper analysis of 'sustainability' is required, which in fact is another motivation for studying how the world could survive for 100,000 years. A society that can do so has truly reached 'sustainability'.

An elementary observation on the state of a society that lasts 100,000 years may be made (significant annual growth is excluded). If the number of people on Earth or their energy consumption doubled every century, an increase by a factor of a million would have already occurred after only 2,000 years. This is absurd: 10 people per square meter and an energy consumption 100 times the energy that reaches us from the Sun. Even an increase of a factor of 10 over the 100,000 years would correspond to an average annual growth rate of no more than 0.003% per year. Hence, such a long-lived society would have to be largely static in its population and its resource use.

There is nothing new in this. In fact, the power of exponential growth was illustrated in an ancient Persian story [7]. Someone had done a meritorious deed and the Emperor asked him what his reward should be. He answered by presenting a chess board, asking that he be given one grain of rice on the first square, two on the second, four on the third and, on every successive square, twice the preceding one until the 64th square was reached. The courtiers laughed about the fool who wanted some rice instead of gold. But when the 20th square was filled, there were already more than a million grains of rice. By the 64th square there would have been 1.8×10^{19} grains of rice on the chess board, corresponding to 30 tons of rice for every man, woman and child in the Earth's population today!

In 1798 Thomas Malthus wrote his famous work *An Essay on the Principle of Population* [8]. He was ridiculed or attacked, though the inescapable truth remains valid: if the population of the world doubled every century (it actually nearly quadrupled during the 20th, see Figure 1.1), there would not even be standing room left on Earth after 20 centuries. Unfortunately various religious

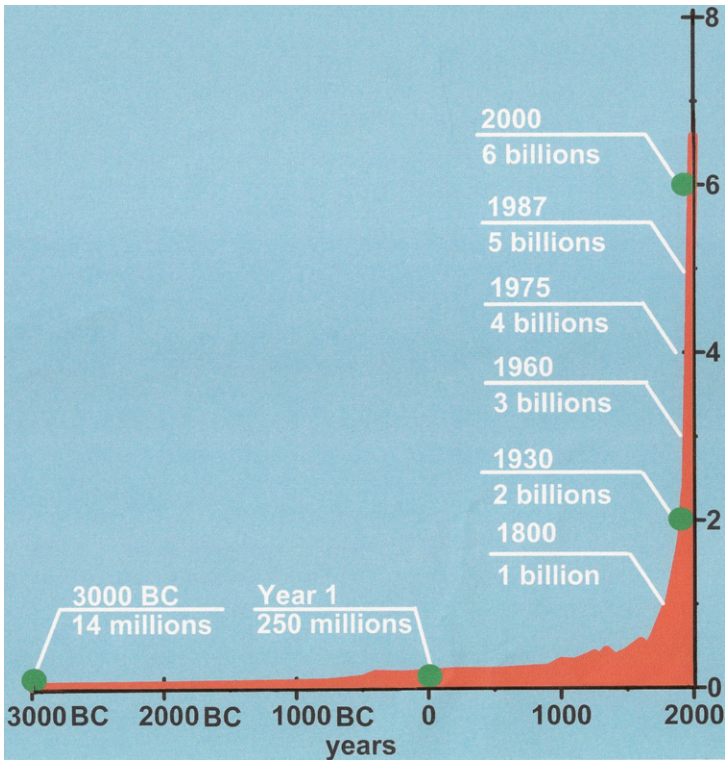


Figure 1.1 The spectacular rise of the human population

organizations, and in particular the Roman Catholic Church, do not wish to acknowledge this truth and thereby are responsible for much human suffering.

Skipping various books and pamphlets published following Malthus, we come to the well-known study, *Limits to Growth*, published under the sponsorship of the 'Club de Rome' – an influential body of private individuals [7]. A first attempt was made to make a complete systems analysis of the rapidly growing human–biological–resource–pollution system. In this analysis the manifold interactions between the different parts were explicitly taken into account. The conclusion was that disaster was waiting around the corner in a few decades because of resource exhaustion, pollution and other factors. Now, 35 years later, our world still exists, and as documented, for example, in Bjørn Lomborg's fact-filled, controversial and tendentious book *The Skeptical Environmentalist*, many things have become better rather than worse [9]. So the 'growth lobby' has laughed and proclaimed that *Limits to Growth* and, by extension, the environmental movements may be forgotten. This entirely misses the point. Certainly the timescale of the problems was underestimated in *Limits to Growth*, giving us a little more time than we thought. Moreover, during the last three decades a variety of national or collaborative international measures have been taken that have forced reductions in pollution,

as we shall discuss. A shining example of this is the Montreal Protocol (1987) that limited the industrial production of fluorocarbons that damage the ozone layer and generated the 'ozone hole' over Antarctica [10]. The publication of *Limits to Growth* has greatly contributed towards creating the general willingness of governments to consider such issues. Technological developments have also led to improvements in the efficiency of the use of energy and other resources, but, most importantly, the warnings from Malthus onward have finally had their effect as may be seen from the population-limiting policies followed by China and, more hesitantly, by India. Without such policies all other efforts would be in vain. However, the basic message of *Limits to Growth*, that exponential growth of our world civilization cannot continue very long and that a very careful management of the planet is needed, remains as valid as ever.

1.2 People and resources

In evaluating the long-term needs of the world it is vital to know how many people there will be and the standard of living they will have. There is, of course, much uncertainty about the level at which the population will stabilize, but in a long-term scenario it cannot continue to grow significantly, though fluctuations are possible as a result, for example, of new diseases. On the basis of current trends the United Nations in 1998 projected as a medium estimate that the world population would attain 9.4 billion in 2050 and 10.4 billion in 2100 to stabilize at just under 11 billion by 2200 [11]. The 2004 revision reduced the 2050 projections to a slightly more favorable 9.1 billion [12]. A probabilistic study in 1997 concluded that in 2100 the most likely value would be 10.7 billion, with a 60% probability that the number would be in the interval 9.1–12.6 billion [13]. Such estimates are based on past experience about the demographic transition from high birth/death rates to low ones. In addition to economic factors, cultural and religious factors play a major role in this, but are difficult to evaluate with confidence. Perhaps somewhat optimistically we shall assume in this book that the world population will, in the long term, stabilize at 11 billion. Our further discussion will show that a number significantly in excess of this risks a serious deterioration in the conditions on Earth. It is, of course, also very possible that instead of reaching a certain plateau the population will fluctuate between higher and lower values. If the amplitude were large, this might be rather disruptive.

Many estimates have been made of per capita consumption levels during the 21st century. Such projections have been based on extrapolating current trends. Since population growth occurs mainly in the less-developed countries, while consumption is strongest in the industrialized part of the world, the implication tends to be that the present level of inequality in the world will continue, with critical resources moving from the former to the latter. Apart from the moral issues, it is not at all evident that this will be feasible as the political equilibriums shift and large countries like China become capable of defending themselves and claiming their share of the world's resources. In any case, for a long-term stable

world to be possible, a certain level of equality between countries might well be required, so we shall adopt here a scenario in which the long-term population of 11 billion people will live at an average consumption level that is comfortable. We shall take this level to be midway between the current level of the most advanced countries in Western Europe and that of the USA. These seem to be relatively satisfactory to their citizens. At constant efficiency the energy consumption would be nearly seven times its present value – a population increase of 1.65 and a per capita increase in energy 4.1 times the present average because of the larger consumption in the less-developed world. If such a scenario is viable, it might be a worthy aim to strive for and orient our future policies in its direction. Evidently, much less favorable scenarios than this 'utopia' can easily be imagined. We also stress that what we are considering here is the viability of our utopic scenario in relation to physical limits. The sociological probability of such a benign future is still another matter. We could, of course, have chosen the US level of energy use as a sole reference and have thereby increased the requirements by 25%. However, since the USA has never considered energy economy a priority, much energy is wasted, and bringing the whole world to the US level would unnecessarily complicate the issue.

In our long-term scenario the fossil fuels would no longer be available. We may be uncertain whether oil and gas will or will not be exhausted within a century, but little would be left after 100,000 years. Only renewable and fusion energy sources could power such a society. Renewables with an adequate yield could include solar and wind energy. Hydropower would remain of interest, but its part would globally remain rather minor. Nuclear energy could make a contribution but with serious associated problems, while renewable biofuels would be in competition with agriculture.

Mineral resources are an essential item in a long-term scenario, since they are not renewable. Recycling should play an increasing role, but can never cover 100% of requirements since losses cannot be completely avoided. Therefore, as Darwin predicted, increasing amounts of energy will be needed to extract metals and other elements from poorer or more inaccessible ores. Several elements may be obtained from sea water, but significant technological development will be needed to minimize the required energy. At the present time, since ores of rather high quality are still available, the motivation for developing extraction technologies for much poorer resources is still lacking.

Fresh water is an essential commodity, in particular for agriculture. Global availability is not an issue, but the uneven distribution is. A significant part of humanity experiences shortages of clean water, although the desalination of sea water could provide clean water in some of the areas in which it is lacking. Of course, desalination takes energy, but, seen globally, the requirements do not seem too problematic. In fact, all the water currently consumed on Earth could be produced by desalination at an energy cost equal to about 11% of present world energy use [14] or even less with newer technologies. With the exception of some desert regions, the present problem of providing many people with clean drinking water is a question of piping, not of availability.

Intensive agriculture is likely to feed the 11 billion people of our scenario a very satisfactory diet, but it requires adequate water, fertilizer and soil. Soil formation is a slow process, and so careful management is needed not to waste what nature has built up during hundreds of centuries. Perhaps fertilizers could become a critical issue, as in nature phosphate availability is frequently a limiting factor to biological productivity [15]. In agriculture, the same may happen in the long term and energy requirements to exploit poorer phosphate resources will become greater. However, current phosphate use is quite wasteful and has led to ecological problems by agricultural runoff, polluting lakes and coastal waters. Hence, a more efficient use would be beneficial to all.

The overall conclusion that we elaborate in Chapter 7 is that a society like our own can survive at a high level of comfort for at least 100,000 years because sufficient renewable energy is available. The problem is not so much the long-term future, but the transition phase from the present state to the distant future. As long as fossil fuels are rather abundantly available, the economic motivation for switching to renewable energy sources is relatively weak. But if we do not begin to do so very soon, the transition may be so abrupt that acute shortages could develop which will be difficult to manage politically in a cooperative framework. Some of these problems are already beginning to be seen today.

The current use of fossil fuels has a major impact on climate, due to the production of CO_2 and other greenhouse gases. The mean temperature of the Earth's surface has risen almost 1 degree Celsius over the last century. Models typically predict some 3 degrees Celsius by the year 2100, and also show that there is much inertia in the climate system. Even if we were to turn off the production of CO_2, it would take centuries to re-establish the previous 'normal' climate and some changes, like the possible melting of the Greenland ice cap, would be irreversible [16].

In the past, when conditions deteriorated, people would simply move elsewhere. But the high population densities of the present world have made this very difficult. Even if global warming made life better in Siberia, this can hardly solve the problem for the more than 2 billion inhabitants of China and India. Moreover, the speed of the changes is so great that it is difficult for nature or people to adjust quickly enough. The increasing warmth causes the sea level to rise – partly because water expands when becoming warmer and partly by the melting of the ice caps. On the 100,000-year timescale large areas of low-lying land might be flooded, displacing millions of people. Such risks show the importance of a rapid switch from hydrocarbons to less dangerous energy sources (see Chapter 6).

1.3 Management and cooperation

The production of CO_2 increases the temperature of our planet. It is immaterial whether it is produced in the USA or in China – the effect is the same. Shortages of phosphate fertilizers may develop. If some countries take more than their

share, others will have less. Most rivers flow through several countries; if one upstream takes all, nothing is left for those downstream. Many other examples could be added. They all have something in common: they can be solved cooperatively and internationally, or will result in economic or military warfare. Inside most countries there are mechanisms to resolve such conflicts. If I and my neighbor have a conflict about water rights, there are national courts of justice whose verdict we are both obliged to respect. If need be there is the national government to enforce these verdicts. As the environmental consciousness develops, the laws under which the courts operate become more directed towards environmental soundness and not just to property rights in the narrow sense. So, in several countries, if I cut down my tree I still need permission to do so or I have the obligation to plant another one.

International environmental treaties and laws are still in a very primitive state, though some successes have been achieved. Probably the Montreal Protocol is the shining example. It took only a small number of years after the discovery that the ozone hole was caused by man-made fluorocarbons to conclude an international agreement to limit or eliminate their production [10]. At the same time it shows that rapid action may be needed. The ozone hole over Antarctica reached its maximum size so far in 2005, but is not expected to regain its coverage before a half century from now.

Another example is the Kyoto convention to limit CO_2 emissions [17]. The convention aiming at very modest reductions in CO_2 emissions was adopted in 1992 at Kyoto and went into effect in 2004. Unfortunately, the USA, the largest producer of CO_2, decided to opt out of the convention as the reduction was too constraining for its industrial interests. Even worse, the USA has tried to line up opposition against it. In fact, the history of the Kyoto convention shows that in the 1990s the general climate for international environmental action was much more favorable than it is today.

Many may think it foolhardy to attempt to project the future for such a long time as a century, let alone 100,000 years. After all, if in 1900 someone had made predictions on the state of the world today, the outcome would have shown little correspondence with the actual situation. But the difference is that today we seem to have a fair knowledge of the physics, chemistry and even of the biology of the world around us. New discoveries in particle physics and astrophysics may be very interesting, but they are hardly likely to change the constraints under which humanity will have to live on Earth. New unforeseen technological developments may lead to undreamed of gadgets that will make many processes more efficient, but they are going to change neither the quantity of solar energy that falls on Earth nor the amounts of metals in the Earth's crust and oceans. So we have a fair idea about the possibilities, unless some entirely unanticipated possibilities appear, like the direct conversion of matter into electrical energy or the creation of totally new forms of life. We also emphasize again that our discussion assumes a 'reasonable' behavior of the Earth's inhabitants, which past experience makes far from obvious. We shall come back to these issues in the last chapter.

1.4 The overall plan of the book

In Chapters 2–5 we provide a general scientific background, beginning with the evolution of the Earth and of life. The Earth was constructed out of a large number of smaller planetesimals with a Mars-size body striking at a late stage. This event created the Moon. It was a fateful and probably not a very probable event, but it stabilized the orientation of the Earth's axis and thereby assured a climate without excessive instability. Smaller bodies, the asteroids and comets, have survived until today. These catastrophic impacts have had a profound influence on the evolution of life. At times whole families of animals were eliminated, thereby creating the ecological space for new evolutionary developments. During the next 100,000 years some such randomly occurring events could be quite destructive to human society, but with proper preparation most could be avoided.

After the formation of the Earth its internal heat began to drive the slow internal flows which moved the continents to and fro and caused some of the crust to be dragged down to greater depth. Cracking of the crustal pieces created Earthquakes causing much regional destruction. Hot liquids from below led to volcanic eruptions sometimes of gigantic proportions. In the process CO_2 was recycled, preventing it from becoming locked up permanently in carbonates. On Earth these subterranean processes and the development of life have ensured relatively stable concentrations of CO_2 and oxygen in the atmosphere, which was beneficial to the further evolution of life. Nevertheless, glacial periods have occurred that put life under stress, and this stress was further amplified when human hunters devastated several ecosystems. The general conclusion from life's evolution is that it was able to respond to slow changes in circumstances, but that very rapid changes were more difficult to cope with. In particular the volcanic mega eruptions and cometary impacts pose much risk to human welfare. For the moment both are difficult to predict.

Current climate change occurs owing to natural causes, but even more due to the production of gases like CO_2 which are enhancing the natural greenhouse effects. Of course we cannot just study the future without looking at the past. Without knowledge of the Earth's history to guide us, we would also remain ignorant about the future. So we shall study what we have learned so far and try to see how things may develop. Past climates contain many lessons for the modeling of future climatic change. As we go back further in time, data about past climates become more limited, but the combination with climate models has made it possible to make some global inferences. Past climate variations, as discussed in Chapter 5, have been mostly related to small changes in the orbit of the Earth around the Sun, to variations in the solar radiance, to volcanic eruptions, to continental displacements, to the coming and going of mountain chains, to changes in land cover, and to variations in the concentrations of greenhouse gases.

Models that successfully account for past climates may also be used to predict future climates and the human influences thereon, as discussed in Chapter 6.

However, the human factor has taken the greenhouse gas concentrations so far beyond the range experienced in the last several million years that uncertainty by a factor of 2 remains in quantitative predictions of the resulting temperature increase. This makes it difficult to foresee whether the Greenland ice cap and the West Antarctic ice sheet will melt with a possible increase in sea level by more than 13 meters and the flooding of much land.

In Chapter 7 we discuss future energy production and mineral resource availability. While the prospects of the 100,000-year society look rather good as far as energy is concerned, shortages of a number of elements will develop, and a significant technological development will be needed to find suitable substitutes. Water, agriculture and forests are considered in Chapter 8, which shows that much care will be needed not to pollute the environment; but if the population is stabilized at 11 billion, adequate food and water can be available and some natural areas may be preserved.

In Chapter 9 we discuss the possibility of colonizing other planets such as Mars and, less likely, Venus. We have analyzed the processes that have made them uninhabitable today to see if these can be reversed, and have concluded that it would be much less difficult to preserve the environment on Earth than to create an appropriate environment on Mars or Venus during the 100,000-year future that we have adopted. We also consider the possibility of extracting resources from the Moon or the asteroids, and again conclude that in realistic scenarios the prospects do not compare favorably with what can be done on Earth.

The primordial importance of international collaborative efforts to ascertain the physical state of the world, and to agree on measures needed to deal with dangerous developments, is stressed in Chapters 10 and 11. Continuous observation of the Earth from space will be needed to monitor the Earth's surface and atmosphere in great detail. Meteorological satellites have shown the benefits of such observations, as have satellites observing the Sun and the Earth's land cover. More is needed and the continuity of the observations and their calibration has to be assured. Not only the instruments are needed but large numbers of researchers to analyze and interpret the data in detail.

Once all the required data are being obtained and analyzed in an agreed fashion, the more difficult problem becomes to actively manage the planet in an equitable manner. It is already clear that the world's CO_2 output is too high and mandatory measures will have to be agreed upon. In other areas of planetary management binding targets may also be needed. The United Nations provides the only existing organization to do so. Many complaints have been made about supposed inefficiencies at the UN, but replacing it by another organization hardly seems to be the solution. Certainly, improvements may be made, but a comparison of the magnitude of its spending with that of the military spending in its member countries shows the unfairness of much of the criticism.

In the final chapter we stress again the need for a firm cooperative framework on managing the Earth on the basis of global data sets. Finally, we turn to the sociological issues: what will be the effect of living in a very different world

without material growth? Will we get bored, or do the examples of more static societies that lasted for millennia, such as Egypt or China, show that another form of society is possible? After all the current growth oriented 'Western model' is itself only some centuries old. Many questions remain unanswered:

- What is the role of religion in a 100,000-year world?
- Is democracy possible?
- Can the arts flower for 100,000 years?

We shall not know until we arrive there, but our more reliable conclusion is that the material basis for such a long-lived society is in all probability assured if humanity manages the Earth wisely. Humanity will have limited choices. Nevertheless, one choice it is allowed to make is to perdure or, alternatively, to destroy itself.

1.5 Notes and references

[1] McRae, H., 1994, *The World in 2020*, Harper Collins Publ.
[2] Darwin, C.G., 1953, *The Next Million Years*, Double Day.
[3] Mellars, P., 2006, 'Going East: new genetic and archeological perspectives on the modern human colonization of Eurasia', *Science* **313**, 796–800
[4] For example, Kastenberg, W.E. and Gratton, L.J., 1997, *Physics Today*, June, p. 41.
[5] Activity report 1996 of the Swedish Nuclear Fuel and Waste Management Company, SKB.
[6] Brundtland Report, 1987, *Our Common Future*, Oxford University Press. (The World Commission on Environment and Development for the General Assembly of the United Nations.)
[7] Meadows, D.H. et al., 1972, *Limits to Growth*, Potomac Associates, London.
[8] Malthus, T., 1798, *An Essay on the Principle of Population*, Penguin Books.
[9] Lomborg, B., 2001, *The Skeptical Environmentalist*, Cambridge University Press.
[10] http://www.unep.org/ozone/treaties.htm
[11] UN, 1998, *World Population Projections to 2150*, United Nations, New York.
[12] UN, 2004, *World Population Prospects: The 2004 Revision*, http://esa.un.org/unpp
[13] Lutz, W. et al., 1997, 'Doubling of world population unlikely', *Nature* **387**, 803–804.
[14] See note 1095 in Lomborg (above).
[15] Emsley, J., 2001, *Nature's Building Blocks*, Oxford University Press, p. 315.
[16] Wigley, T.M.L., 2005, 'The Climate Change Commitment', *Science* **307**, 1766–1769.
[17] UN, 1992, *United Nations Framework Convention on Climate Change*, UNFCCC, http://www.unfccc.int/

2 A Brief History of the Earth

The universe we observe has precisely the properties we should expect if there is, at bottom, no design, no purpose, no evil, no good, nothing but blind, pitiless indifference.

Charles Robert Darwin

Compared to its sister planets in the Solar System, the Earth is the most complex of all. It could be called a 'complete planet'. It possesses everything that we find individually in the other planets: impact craters, volcanoes (Venus, the Moon, and Mars), a magnetic field (Jupiter, Saturn, Uranus, Neptune ...), a moon and an atmosphere. All other planets are simpler than the Earth as they evidence only a few of these features. That completeness makes the Earth extraordinary, were it not for the single fact that it shelters life due to the presence of abundant liquid water. The Earth is a system, and all its components interact with each other in a way that makes its understanding more difficult. In view of discussing the possibilities of settling human colonies on Venus and on Mars, as we shall do in Chapter 9, it is important to recall how the Earth was formed, how it evolved and how life originated on it.

2.1 The age of the Earth

The age of the Universe is now fairly well established at 13.7 billion years from concordant observations made with the Hubble Space Telescope and the WMAP mission of NASA which has measured, with a precision of a few parts per million, the surface brightness fluctuations of the first light emitted after the Big Bang. Projecting the next 100,000 years is only guessing about a minute fraction of time less than 10^{-5} of the age of the Universe. On our relative timescale, where the age of the Earth, established at 4.56 billion years, is set equal to the duration of a terrestrial day, our exercise deals with less than 2 seconds after midnight! In a purely astrophysical context, this should not be too risky! Why should we worry then?

The problem is that the Earth is not a dead body; it is evolving on relatively short times as the consequences of physical phenomena – orbital variations, cosmic bombardment, natural geophysical phenomena – and also because of environmental modifications resulting from the presence of life, and of humans in particular who have been able to develop technologies leading to a population

explosion and the domination of the entire planet. It is, in every sense of the words, a 'living planet', suggesting also that it may come to an end! This is indeed foreseen when the Sun becomes a red giant and swallows the Earth – a topic discussed in the next chapter. But how do we know so precisely the present age of the Earth?

The material we dispose of is what remains of the building blocks that made up the Solar System: the entire solid bodies such as asteroids and comet debris, interplanetary dust and, of course, the solid planets and their satellites, among which our own Moon offers one of the most precious tools for studying the early history of the Earth and of the whole Solar System. Mars also can serve as a comparative tracer of the evolution of the early Earth where erosion processes and plate tectonics, which were less efficient or non-existent on Mars, have erased the traces of its young ages.

At several places in this book we refer to techniques for dating past geological or climatic events using a special type of clock called radioactive decay, which is based on the properties of certain isotopes of a given atom to decay and form isotopes of another species. The rate of decay of a radioactive isotope is exponential and measured by its half-life, the time at which a large number of atoms have decayed to half the original number.

Table 2.1 gives some examples of various parent and daughter isotope(s), together with their evaluated half-lives. The determination of the half-life is made in the laboratory and can be measured in days or years. Some radioactive isotopes decay much more slowly, and as their half-life can extend to several millions or billions of years they are the most appropriate for dating the ages of meteorites or old rocks. The longer the half-life, the more delicate is its determination, however. The heaviest parent isotopes were synthesized in the explosions of massive stars that scattered materials through the galaxy, out of which the stars and their planets were eventually formed.

Table 2.1 Examples of radioactive elements and their daughter isotopes, together with their agreed half-lives

Parent isotope	Stable daughter isotope	Half-life (years)
Uranium-238	Lead-206	4.5 billion
Uranium-235	Lead-207	704 million
Thorium-232	Lead-208	14.0 billion
Lead-202	Lead-204, 206, 207, 208	53,000
Rubidium-87	Strontium-87	48.8 billion
Potassium-40	Calcium-40	1.4 billion
Potassium-40	Argon-40	1.25 billion
Samarium-147	Neodymium-143	106 billion
Hafnium-182	Tungsten-182	9 million
Aluminum-26	Magnesium-26	700,000
Carbon-14*	Nitrogen-14	5,730

*See Box 2.1

Box 2.1 Carbon-14 dating

One of the best-known dating methods is through the carbon-14 (^{14}C) radiometric technique. With a half-life of 5,730 years, ^{14}C should now be extinct because it has been decaying since the beginning. Therefore, ^{14}C dating is of no use for dating old rocks and meteorites. However, ^{14}C is continuously created through collisions of cosmic rays with nitrogen in the upper atmosphere, ending up as a trace component in atmospheric CO_2. Living organisms absorb carbon from CO_2 through photosynthesis for plants and consumption of living organisms for animals. After death, no new ^{14}C is ingested in the organism and whatever amount is present therein decays with the half-life of 5,730 years. Hence, the proportion of carbon-14 left in a dead organism at a given time allows us to date its death. ^{14}C dating accuracy is limited to around 58,000 to 62,000 years. This accuracy is affected by various other natural phenomena such as local volcanic eruptions which release large amounts of CO_2 in the atmosphere, solar wind and the modulation of the incoming flux of cosmic rays by the interplanetary and the geomagnetic field, or anthropogenic/industrial activities and atmospheric nuclear explosions. Such perturbations require accurate, careful and multiple cross-calibrations.

Detailed descriptions of radioactive dating are the object of an abundant literature [1]. The precision has been increased as more and more accurate measurements of half-lives became available. Ambiguities exist, however, especially when it comes to the point of estimating the original amount in the parent and in the daughter isotope. The presence of several isotopes, in particular those that are not formed by radioactive decay, allows these ambiguities to be corrected. This is the case of strontium-87 (^{87}Sr), the product of rubidium-87 (^{87}Rb) with a half-life of 49 billion years, which has another stable isotope, ^{86}Sr. This is also the case of the lead isotopes: lead-204, 206, 207 and 208, which proves very useful for a variety of materials such as igneous rocks, sediments and ice cores.

Dating through uranium–lead decay is one of the oldest methods, with accuracies reaching better than 2 million years for rocks about 3 billion years old. Uranium–lead dating has been used for the dating of zircons, which are crystals formed from the Earth's magma. They are made of zirconium (hence their name), silicon and oxygen ($ZrSiO_4$), with traces of other elements including uranium. They are originally lead-free when they form because lead atoms are too large to be incorporated in the lattice of the crystal. The radioactive decay of uranium-235 into lead-207 with a half-life of about 700 million years, and of uranium-238 into lead-206 with a half-life of about 4.5 billion years, allows an accurate determination of the age of the sample to within 0.1%!

Using this technique, zircons have been discovered in the Jack Hills and

Mount Narryer regions of Western Australia with a crystallization age of 4.4 billion years before present (BP) [2], making them the most ancient minerals so far identified on Earth. In contrast, the oldest rocks (Acasta gneiss from northwest Canada) date to 4.05 billion years BP [3]. The oldest material of condensates in our Solar System are the Calcium–Aluminum-rich Inclusions (CAI) [1] found in primitive meteorites which have an age of 4.56 to 4.57 billion years [4]. This age is also considered to be the age of the Sun and of the whole Solar System.

What happened on the Earth during the (approximately) 150 million years between the condensation of the CAI and the formation of the zircons is largely unknown. Current models show that the heat generated by the accretion of planetesimals, plus the radioactive decay – at that time six times more intense than it is today - and the liberation of gravitational energy, were enough to keep the entire planet molten. This allowed the metals such as iron and nickel (32% of the Earth's mass) to separate from the lighter silicates and migrate to the center, leaving behind a mantle of primarily silicates. The iron–nickel core was formed very rapidly in the first 30 million years as may be seen from the abundances of a tungsten isotope (see Box 2.2 [5]). The core consists of a molten outer core and a solid inner core.

The early formation of zircons on the Earth seems to indicate that at least some of the crust at the surface had solidified before then. When zircons crystallize, titanium oxide (TiO_2) is incorporated into their lattice, depending on the crystallization temperature. The solidification temperature of the molten magma, on the other hand, depends on its water content. It turns out that most of the Jack Hills zircons crystallized at around 700°C, which is about the same as for modern-day igneous zircons [6]. Since dry rock would have a melting temperature several 100°C higher, this indicates that liquid water was present near the surface of the Earth at the time of formation of these zircons. These low temperatures are confirmed by $^{18}O/^{16}O$ data: both ^{16}O and ^{18}O are stable isotopes of oxygen and their proportions in a crystal depend on the temperature at the time of the crystal formation and can be used as a thermometer (see Box 5.3, 'Information from isotopic abundances' on page 164). Together with the granitic composition of some of the inclusions in the zircons, this would tend to indicate that the present cycle of crust formation, erosion and sediment recycling was already established within some 100–150 million years after the formation of the Earth and that also some early oceans could already have existed [2].

2.2 Geological timescales

Figure 2.1 represents on a linear scale the duration measured in billions of years BP of the geological epochs that marked the Earth's history. The names given to these eons date back to before the application of radioactive dating. Absolute ages are now available through radioactive dating and have become increasingly

Box 2.2 When did the Earth's core form?

Tungsten-182, or ^{182}W, was present in the material from which the Solar System formed as seen from meteoritic (chondritic) samples. It also is produced in the decay of radioactive hafnium, ^{182}Hf, with a half-life of 9 million years. Hafnium is a lithophile element, which means that it tends to be preferentially dissolved in silicates. Tungsten is siderophile, which means that it preferentially dissolves in iron. When the Earth's core formed, most of the tungsten went into the core and, as a result, the hafnium/tungsten ratio in the mantle became larger. If the core formed late, let us say after 100 million years, the radioactive ^{182}Hf would have decayed to ^{182}W and most of the latter would have joined the other tungsten isotopes in the core. But if the core formed much earlier, the ^{182}Hf would still have been around and stayed in the mantle to decay at a later stage, in which case there would be an excess of ^{182}W in the mantle compared to the total W. By measuring the fraction of ^{182}W we can therefore determine how much of the radioactive hafnium originally present at the formation of the Solar System had decayed, and therefore how many half-lives of ^{182}Hf had passed at the time of core formation. The isotope ratio in tungsten changed by about one part in 10,000, and only recently has the instrumental sensitivity become adequate to detect such small effects. Three recent sets of measurement now agree and the conclusion is that the core formed, at most, 30 million years after the formation of the Earth.

accurate to better than a few million years. The period in which the presence of life was unambiguously discovered in the form of macrofossils is called the Cambrian period and goes back about 540 million years. The time before this was simply called the Precambrian, which was later divided into the Hadean, the Archean and the Proterozoic. The Hadean eon, between 4.5 and 3.8 billion years, corresponds to the phase of the formation and accretion of the Earth and its differentiation. It also includes the Late Heavy Bombardment period described in the following section. The Archean eon extends between 3.8 and 2.5 billion years BP. The oldest rocks exposed on the surface of the Earth today are from that period. The atmosphere was low on free oxygen and the temperatures had dropped to modern levels.

The Proterozoic eon starts about 2.5 billion years ago with the rise of free oxygen in the atmosphere. The first unambiguous fossils date back to this time. The late Proterozoic (also called the Neo-Proterozoic) terminates with the brief appearance of the abundant Ediacaran fauna (Section 2.6). The whole period, starting at 543 million years BP, is referred to as the Phanerozoic; it is marked by the rapid and extreme diversification of life (the so-called Cambrian explosion). During the Phanerozoic the geological periods have generally been defined by the presence of characteristic fossils. The separations between the periods frequently correspond to major extinctions. The principal eras of the Phanerozoic are the

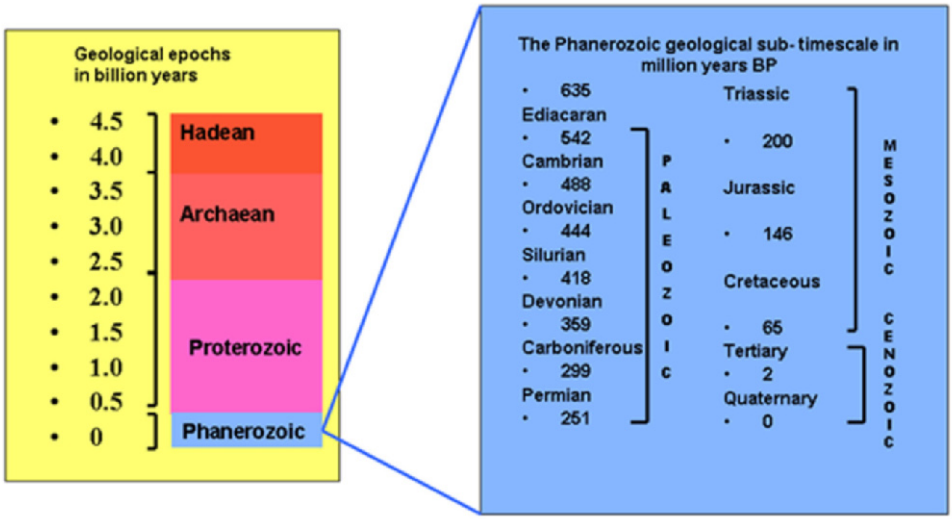


Figure 2.1 Linear scale duration of the geological epochs measured in billions of years BP, with the various eons marking the Earth's history indicated in color for each epoch. The right part of the figure is an enlargement of the Phanerozoic sub-timescale measured in millions of years BP.

Palaeozoic (542–251 million years, ending with perhaps the largest mass extinction of life), the Mesozoic (251–65 million years, ending with the extinction of the dinosaurs and many others) and the Cenozoic (65 million years to present), which is divided into the Tertiary (65–2 million years) and the Quaternary. The latter is divided into the epochs Pleistocene (2 million years to ~10,000 years) and the Holocene, the period after the last ice age during which humans had settled throughout most of the Earth. The two extinction events at the transitions between the Phanerozoic eras are frequently referred to as the Permian–Triassic (P–T) and the Cretaceous–Tertiary (K–T) extinctions, respectively, the K referring to 'Kreide', the German name for Cretaceous. The current eras and periods of the Phanerozoic era and the ages of their separations in million of years are as indicated on the right part of Figure 2.1.

2.3 The formation of the Moon and the Late Heavy Bombardment

The scenario for the formation of the Solar System is now well supported by the observations of by the Hubble Space Telescope of planetary systems, giving strong support to the model of a proto-solar nebula that results from the gravitational collapse of a cloud of interstellar dust and molecular gas, and forms rings of denser concentrations of dust and of small proto-planets or planetesimals. It is now accepted that about 500 such bodies of approximately the size of the Moon have accreted and formed the inner planets of the Solar System [7].

The formation of the Sun and the planets through accretion did not stop abruptly but continued with decreasing intensity. The scars of this natural bombardment are the craters that are observed on the surface of all the solid bodies of the Solar System. Their numbers per unit area are used as a tool for the relative dating of their respective parent bodies. Some of these bodies, however, are affected by erosion processes, plate tectonics and volcanism. Objects with atmospheres, such as the Earth, Venus and Titan, have a bias towards the larger bodies hitting their surface. Since small impactors burn up in the atmosphere they do not reach the surface and consequently there is an obvious predominance of large impact craters of tens and hundreds of kilometers in diameter on these bodies. In addition, Venusian volcanism, as observed by NASA's Magellan radar imaging mission, has erased any evidence of early craters. Without such volcanic activity, it would have been possible to reconstruct the history of Venus's atmosphere and understand better why and how it reached a monstrous pressure of 92 bars of CO_2, 100 times the Earth's atmospheric pressure.

The Moon, Mercury and Mars, which have no or just a low-density atmosphere, do offer a coherent crater sample with some similar properties. On the Moon, the most ancient terrains present the highest level of cratering, the highlands, while the Mare are smoother and younger. The Aitken basin at the lunar South Pole, roughly 2,500 km in diameter and 13 km deep, is the largest known impact crater in the entire Solar System [8]. The material ejected from the lunar soil most probably comes from the mantle of the Moon, as evidenced by chemical composition anomalies observed by the American Clementine satellite and by gamma-x-ray spectroscopy. The Moon is particularly interesting because the samples brought back by the Apollo astronauts and by the Soviet robotic Luna missions provide an absolute scale for dating the meteoritic bombardment through isotopic analysis. These samples have revealed the chemical and mineral compositions of the Moon's soil [9]. Since we do not have corresponding samples from other objects, the dating of the surface ages of these bodies in the Solar System, in particular of Mars, comes essentially from the dating of lunar rocks.

These samples show some similarities between the compositions of the Earth and the Moon, but are depleted in volatile elements with respect to the Earth mantle, such as potassium (Table 2.2). The primitive Earth mantle and bulk Moon composition are chemically similar, while volatiles such as C, K and Na are less abundant in the bulk Moon than in the primitive Earth mantle. This observation is compatible with what would result had the Earth been hit by a large body more or less the size of planet Mars [5], ejecting into space a large fraction of the Earth's mantle, with little iron because iron was already differentiated into the Earth's core. In the course of this huge impact, the refractory part of the hot disk formed by the ejected debris condensed and eventually formed the Moon. Part of the material from the impactor merged into the Earth, but this material must not have had a composition that was too different from that of the Earth since the Moon's composition is more or less the same as that of an Earth mantle just deprived of its most volatile elements. This is

Table 2.2 Comparison between the chemical compositions in % by weight for different types of rocks on the Earth (adapted from Lodders and Fegley [10]). All values are in % of the weight of the rock. For the definition of crust and mantle, see Table 2.3

Element	C1 meteorite [10]	Earth's primitive mantle*	Moon's bulk [11]	Moon's highland crust [11]
O	46.4	44.4	43.0†	44.0†
Fe	18.2	6.3	10.6	5.1
Si	10.6	21.0	20.3	21.0
Mg	9.7	22.8	19.3	4.1
Al	0.86	2.35	3.2	13.0
Ca	0.93	2.53	3.2	11.3
Na	0.50	0.27	0.06	0.33
K	0.055	0.024	0.008	0.06
C	3.45	0.012	0.001	~0.0001†

* Primitive Earth Mantle = Mantle + Crust + Hydrosphere (see Anderson [12])
† Estimated by J. Geiss.

a strong indication that the impactor came from the same region of the solar nebula as the Earth. The similarity between the composition of the Moon and the Earth implies that the impactor had an orbit close to that of the Earth.

This scenario can explain very well several of the Moon's characteristics. First, the Moon's size is unusually large relative to its mother planet when compared to other natural satellites in our Solar System. This makes gravitational capture highly improbable. Second, if the Moon would have accreted with the Earth, it would circle the Earth in the ecliptic plane. The tilt of its orbit of 0.5 degrees relative to the ecliptic is high and is most likely the result of the Moon-forming impact. Third, as mentioned previously, the Moon is totally devoid of volatile elements such as hydrogen, carbon, nitrogen and potassium. Fourth, the highly differentiated anorthositic lunar crust – made of feldspar, like most of the Earth's crust – formed very early from a magma ocean, and the necessary heat to provide such an ocean could only have been provided through a fast accretion event [13, 14]. This theory has been recently reinforced by the precise dating of our natural satellite to between 25 and 33 million years after the formation of the Solar System and of the Earth (see Box 2.2).

The Moon-forming impact had very important consequences. The Earth–Moon distance was probably no more than 25,000 km and increased rapidly since then to reach its present lunar orbit of 400,000 km. The laser reflectors that have been placed on the Moon by the Apollo and Luna missions show that this distance is still slowly increasing at a rate of 3–4 cm per year. Under the shock, the Earth went into a spin and its axis of rotation tilted, which, after evolution and tidal interactions with the Moon, determined the present cycle of seasons and the duration of our day. The tidal forces were huge, inducing wave motions of the thin and malleable crust as much as 60 meters twice a day [15]. The tight

Table 2.3 Data on the Earth's interior

Thickness (km)		Density (g/cm^3)	
		Top	Bottom
Crust	30	2.2	2.9
Upper mantle	720	3.4	4.4
Lower mantle	2,171	4.4	5.6
Outer core	2,259	9.9	12.2
Inner core	1,221	12.8	13.1
Total	**6,371**		

Source: Anderson [12].

gravitational coupling between the two bodies has also distorted the shapes of both the Earth and the Moon. Eventually the Moon turned only one face towards the Earth. The tidal coupling also stabilized the Earth's rotation axis, protecting our planet against strong climate changes that would have made the evolution of life much more difficult [16].

The Earth and its newly-formed Moon continued to be bombarded as the gravitational attraction of the bigger planets, Jupiter and Saturn, altered and elongated the orbits of debris remaining from the original accretion disk, as they slowly migrated towards their present orbits [17]. These small bodies impacted Mars, the Earth and the Moon, Venus and Mercury, delivering water ice and other frozen volatiles (Section 2.5). Unfortunately there is no evidence of the early meteoritic and cometary bombardment on the Earth because erosion has removed all traces of these events. The Moon, therefore, provides a good record of the early impact history in the inner Solar System. Radioactive decay dating of Apollo samples showed that the bombardment diminished gradually until around 4 billion years BP. However, the large Mare basins that were formed by partial melting due to large impacts on the Moon have younger ages in the range 3.9–3.8 billion years, indicating that there was a sudden increase of large impacts. This 'cataclysm' period has been called the 'Late Heavy Bombardment' (LHB) as it was found that just not one single event but rather a fast succession of impacts occurred between 3.9 and 3.8 billion years and span the best calibrated epoch in lunar history between 4.1 and 3.1 billion years [18]. There is some evidence that the southern hemisphere of Mars also experienced large impacts around the same time, and that the Caloris basin on Mercury may have a similar origin [19]. Until recently, the Earth had hidden any kind of evidence of this event, but in 2002, tungsten isotope anomalies were found in early Archean sediments from Greenland and Canada which indicate the presence of an extraterrestrial component in these rocks, providing a possible 'fingerprint' of the LHB on the Earth [20].

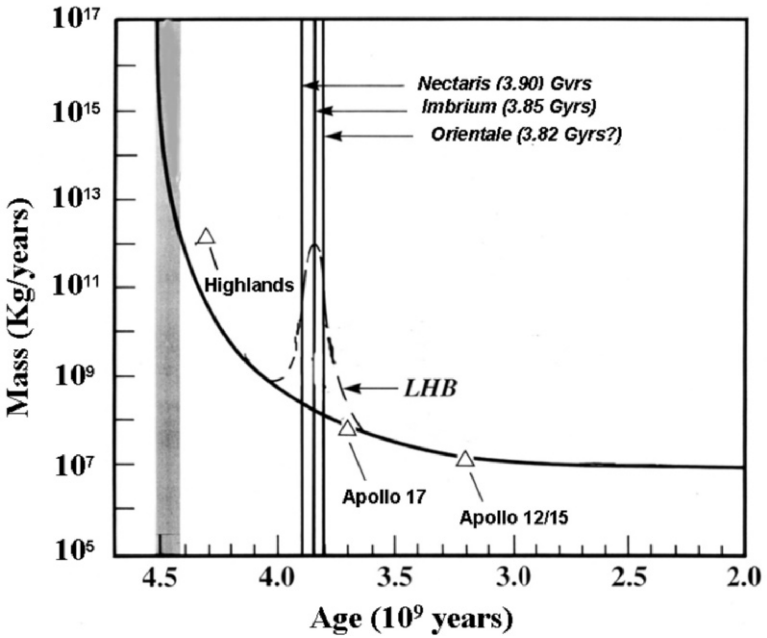


Figure 2.2 Accretion rate in kilograms per year on the Moon. Triangles mark data from lunar Apollo sample studies, and the formation of the lunar highlands. Ages of a few major impact basins are indicated. The solid line is the present-day background flux extrapolated back in time towards the origin of the Solar System. The spike around 3.85 billion years corresponds to the Late Heavy Bombardment. (By permission of C. Koeberl [see reference 22].)

What could have caused the LHB? One possible model suggests the migration of the giant planets. Gravitational resonances could have destabilized the orbits of the volatile-rich objects within a short period of time outside the orbits of Jupiter and Saturn. Some of these objects may have reached the inner Solar System and caused the LHB [21]. At the same time, the outer edges of the asteroid belt were affected by these instabilities, adding another contribution to the bombardment. Today Saturn and Jupiter have cleared the region between 5 and 30 Astronomical Units* of any such objects. The rate of mass accretion during the LHB was a few thousand times larger than any period before or after, showing the exceptional nature of this period. The LHB lasted some 100–200 million years (Figure 2.2. [22]) during which, impactors of various sizes created craters larger than 20 km every 100 years on average, some

* An Astronomical Unit (AU) is equivalent to the mean Sun–Earth distance or about 150 million km.

reaching 5,000 km, as large as South America, strongly modifying the environment of our planet not long before durable life may have appeared (Section 2.6). We do not know whether or not life had already originated during the Hadean, and if it had, it could have survived the LHB. But if we believe the evidence for life in the early Archean, it apparently did not take very long for it to develop after the LHB had ceased.

2.4 Continents and plate tectonics

2.4.1 Continents

As early as the transition between the Hadean and the Archean, the shaping of buoyant continents and of oceanic crust was initiated. The heat produced by both the pressure exerted by the upper layers and by the decay of radioactive material drove convective motions of the hot viscous interior of our planet; the cooler material plunged down, was heated again and rose back up. Convection was very active from the beginning as an efficient mechanism to cool the hot core and the solid mantle, carrying upward interior heat through slow vertical motions of partly melted material. As the crust started to solidify, it formed the first elements of tectonic plates, even though at that time they were not necessarily drifting apart.

At mid-ocean ridges, under water or above isolated hotspots, the dense oceanic crust is constantly destroyed and renewed every 100 million years or so on average. It is mostly constituted of basalts richer in iron (9.6%), magnesium (2.5%) and calcium (7.2%) than the continental crust, which is light and dominated by granites that are richer in silicon (Si, 32.2%) and potassium (K, 3.2%), both containing about 45% of oxygen by mass. These characteristic differences are due to the presence of water which plays a fundamental role in hydrating minerals and forming granites which themselves form the continents [23]. In reality, the formation of granites is very complex and not fully understood. It is probably the result of partial melting of that part of the continental crust that is lying above hotspots in the mantle below, and during subduction of the oceanic crust, because once a solid continental crust has been formed [24, 25], it remains stable and cannot be renewed by subduction (Chapter 4).

The Archean plates were small and pervaded by the extrusion of basaltic material above hotspots which probably formed the first proto-continents. Their granitic core evolved through the melting of basalt above the hotspots. As the heat flow decreased, the area of the plates increased, reaching sizes comparable to the plates of today. After an abrupt increase in crustal volume between 3.2 billion years, when the continental volume was only 10–20% of what it is today, and 2.6 billion years BP, when it reached 60%, the crustal volume continued to increase but less rapidly throughout all the Proterozoic until more recent times (Figure 2.3). Contrary to the ocean floor, these early continents emerged above an ocean-dominated planet and accumulated over billions of years. They are the

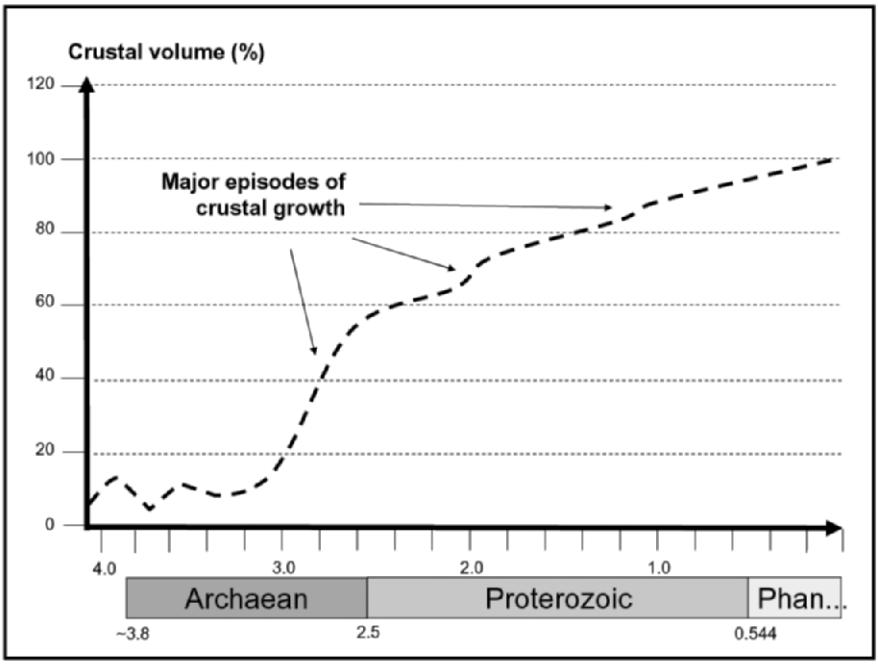


Figure 2.3 Major episodes of crustal growth through the eons. (Adapted from Ashwal [24], and Taylor and McLennan [25].)

source of the records that permit the reconstruction of the past geological and biological history of our planet, despite the difficulties of deciphering their messages due to erosion processes. The tail end of the formation of the continental crust is evidenced today by the motions (mostly horizontal) of tectonic plates.

2.4.2 Plate tectonics

As early as 1596 the Dutch map maker Abraham Ortelius suggested that the Americas, Eurasia and Africa were once joined and have since drifted apart creating the Atlantic Ocean. Several centuries later, this idea raised a genuine interest in the mind of some curious people such as Benjamin Franklin and Alexander von Humboldt. In 1912, Alfred Wegener [26] from Germany revisited of Ortelius's suggestion, watching carefully the close fitting of the west coast of Africa and the south coast of America. He then proposed that these two separate continents were once compressed in one single 'super-continent' that he called Pangaea (Figure 2.4). His idea explained well the continuity of the formation of mountains on both sides of the Atlantic and the fact that more or less the same plants and fossils of the same age are found all around the world today. The theory has since then been developed, explaining and providing a rational basis to these apparently extraordinary coincidences.



Figure 2.4 The Pangaea super-continent existed in the mid to late Permian. Shown in green, are the Precambrian continents or sectors that may have pertained to Rodinia which existed 750 million years ago, as established through reconstructions. (Courtesy of Torsvik [27], by permission of the magazine *Science*.)

It is quite difficult to reconstruct the motion of the plates throughout the Earth's history. At least for the recent past it is easier due to the more precise knowledge of the present velocity and direction of each plate, as provided by accurate satellite measurements. Space-borne observations, in particular geodesy satellites associated with more and more sophisticated models of the interior of the Earth, have indeed added their investigative power in support of the theory. Seismic studies of the waves generated during earthquakes (Chapter 4), together with the studies of the crust – which is easily accessible to us from the ground – of the rocks and, of course, of the fossils as well as the technique of paleomagnetism (see Box 2.3) all contributed to this significant improvement in our understanding of the dynamics of the solid Earth.

The development of the theory of plate tectonics, together with the

Box 2.3 Paleomagnetism

Paleomagnetism refers to the study of the past variations of the orientations of the Earth's magnetic field. One method is based on Thermal Remnant Magnetism. Minerals containing iron-oxides called magnetite, usually found in basalt and other igneous rocks, can act as natural compasses that record the orientation of the field. The Curie point is the temperature at which the spin of the electrons in a mineral are subject to a transition from a free to a 'fixed' orientation and vice versa. For magnetite, this temperature is 580°C, well below the crystallization temperature of the rocks themselves, usually about 800–900°C. As it cools through the Curie point temperature in the presence of a magnetic field, the mineral becomes ferromagnetic and the magnetic moments are partially aligned with the magnetic field. If the Earth's field is fixed and stable, these 'mineral compasses' provide a powerful means of determining any variation in the orientation of a continent and of its drift. Radioactive decay dating (usually potassium–argon and argon–argon) allows us to reconstruct the motions of the portions of the Earth or of continents through the past. However, the field itself suffers reorientations and changes in polarity, and the Earth's rotation axis is also subject to tumbling. Nevertheless, these reorientations or motions result in synchronous changes in latitude all around the globe, while continental shifts are not necessarily synchronous. The basalts forming the seafloor, for example along the mid-Atlantic ridge, do offer the best records of these orientation changes and allow not only the reconstruction of the past history of the field's intrinsic orientations (synchronous) but also of the continental drifts.

Other methods use the orientation of magnetic grains themselves bound to sediments (Depositional Remnant Magnetism) or found in chemical solutions later mineralized (Chemical Remnant Magnetism) like hematite or sandstones.

Paleomagnetism has been very instrumental in verifying the theories of continental drift and plate tectonics in the 1960s and 1970s. Paleomagnetic evidence is used also in constraining possible ages for rocks and processes and in reconstructions of the history of the deformations of parts of the crust.

observations of the different continents, indicates that the Earth's crust is presently broken into about 10 plates, which move in relation to one another, shifting continents, forming new oceanic crust, and stimulating volcanic eruptions. Today's tectonic plates are gigantic, the biggest one having the size of the Pacific Ocean. Their displacements, because of the spherical shape of the Earth, are rotational and measure from 2 to 16 meters per century. Over periods of 10 million years or more, this corresponds to displacements reaching 1,000 km or more. Even over 100,000 years this is not negligible, reaching 2 to 16 km. Very spectacular is the race of the Indian plate, converging towards the Eurasian

plate at the incredible velocity of 15 meters per century, traveling some 4,500 km in less than 30 million years. Today, India is still pushing against the Himalayas at the enormous rate of 2–4 cm per year (2–4 km over 100,000 years).

Tectonic plates include the crust and part of the upper mantle. They are far from homogeneous even though they are rigid because they are cold. Their thickness varies from about 100 km to only a few kilometers underneath the oceans, in the area of the Hawaiian Islands. Overall, plate tectonics theory has profoundly revolutionized our understanding of geophysical and natural phenomena, and has placed Earth sciences on a solid rational ground. These analyses confirmed the suggestion by Wegener of a unique Pangaea continent, at about 250 million years ago. This is apparently not just a single and unique coincidence in the past history of the Earth, as paleomagnetic studies and analyzes of rock relationships suggest that another super-continent, named Rodinia, also existed some 750 million years ago. Reconstructing the continental drifts beyond 1 billion years in the past has been attempted, but is by no means easy due to the lack of reliable records of fossils, of paleomagnetism and of rocks, and also because of the continuous creation of continental land mass which covers the traces of older grounds. As shown in Figure 2.4, Pangaea was formed of parts of Rodinia and Gondwana, which was formed in the late Precambrian, 550 million years ago. The subcontinents Laurentia and Baltica together (with Avalonia in between, not shown in the figure) combined 418 to 400 million years ago to form Laurussia. It is the collision of Gondwana and Laurussia that formed Pangaea [27].

Gondwana started to break up at about 160 million years ago, delineating the known continents of today: Africa and South America, Antarctica and Australia, and then India. At 65 million years, Africa had connected with Eurasia and the Atlantic Ocean opened. It is foreseeable that this spreading will continue for the next 50 million years and beyond while a new sea will open in the east of Africa, and Australia will cross the equator [28]. It is possible that in another 250 million years from now, the plates will once again join together in a new Pangaea. Geologists suspect that this process is cyclic with new Pangaea forming every 500 to 700 million years. This is clearly outside the 100,000-year limit considered in this book.

These continental displacements are obviously affecting the climate and life on Earth, in particular through the modifications the displacements induce in the oceanic circulation and shifts of ice masses. The biomass and life forms are thereby homogenized all around the globe. The regulation of the water and land temperatures undergoes strong modifications whose effects are as important as any other sources of climate change. The big differences, however, are the relative time constants of these changes, in the range of millions of years as compared to anthropogenic modifications that can be effective in a few decades only, and in the course of 100,000 years other natural or anthropogenic causes ought to be considered. This is the subject of Chapters 5 and 6.

2.4.3 The Earth's magnetic field

The existence of a magnetic field has been important in allowing the development of life on Earth. The source of the Earth's magnetic field and its fluctuations is to be found in a dynamo process located at about 3,000 km below the surface. The field is activated by currents in the liquid kernel of iron and nickel that is kept in motion by convection, by Coriolis forces and gravitation, amplifying the remnant of the original field of the solar nebula. The Earth's field can be assimilated to a dipole or to a magnet with a north and a south magnetic pole, with lines of force joining both poles. Figure 2.5 [29] shows that this simple model does not fully represent the complexity of the field, with some of the magnetic lines closing other than at the poles, constituting what is called the non-dipolar field. The Earth's magnetic field sustains the magnetosphere, which acts as a protective shield against the lethal particles of the solar wind, and those that are emitted during solar eruptions and ejections of matter from the solar corona. A plasma torus called the Van Allen radiation belts, discovered by the first American satellite Explorer 1 in 1958 by US physicist James van Allen (hence their name), stores the high-energy particles. The outer belt, which extends from an altitude of about 13,000–65,000 km above the Earth's surface, contains high-energy electrons and various ions (mostly protons) which have been trapped by the magnetosphere (see Figure 3.2 in Chapter 3). The inner belt extends from an altitude of 700–10,000 km above the surface and contains high concentrations of highly energetic protons. They present a genuine danger for technical systems which can be irreversibly damaged when they cross the belts, as well as for living bodies whose genetic material might be transformed or destroyed. This is the reason why the orbits of satellites usually avoid crossing these belts too often.

How does the Earth's dynamo work? By theoretical principles, the mechanism ought to result from complex interactions within the Earth. The formation of the field by a self-excited dynamo is facilitated by instabilities in the field's orientation and its polarity [30, 31]. Laboratory experiments conducted at Ecole Normale Supérieure de Lyon, using liquid sodium in rotating turbines, have succeeded in reproducing most of the characteristics of the Earth's dynamo, and in particular the field reversal [32, 33]. On average, the Earth's north and south magnetic poles have swapped every half-million years over the past 160 million years. Beyond that, the quality of the data is decreasing with time before present. Over the last few million years, an acceleration of the process has been observed with some five well-identified reversals in the last 2 million years (Figure 2.6 [34]). Over the last four centuries, the strength of the field has continuously decreased. A comparison of data obtained by the Danish Oersted satellite in 2000 with those from the American satellite Magsat 20 years earlier, provides evidence of this decline. Accordingly, it is probable that the Earth's field will reverse within the next 100,000 years. No one can predict how long an inversion period might last; it may vary between a few thousand and tens of thousand years.

Hopefully, the field would probably not completely disappear during that inversion but would rather shift from a predominantly bipolar field to a multipolar field, with a large number of local north and south magnetic poles. If

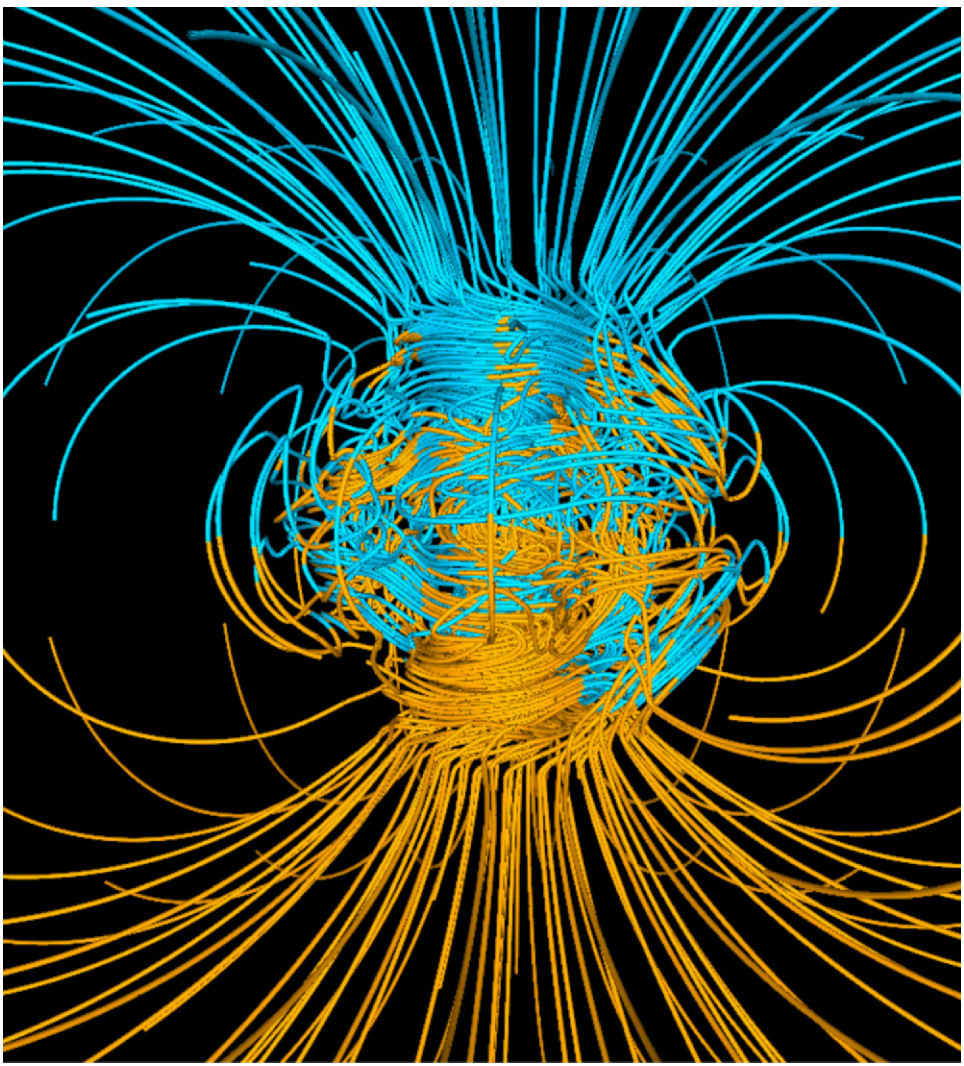


Figure 2.5 The present structure of the Earth's field lines showing the superposition of a north–south dipolar field and of several multipolar components. Yellow and blue represent the two polarities of the field. (*Credit*: G.A. Glatzmaier, reference [29].)

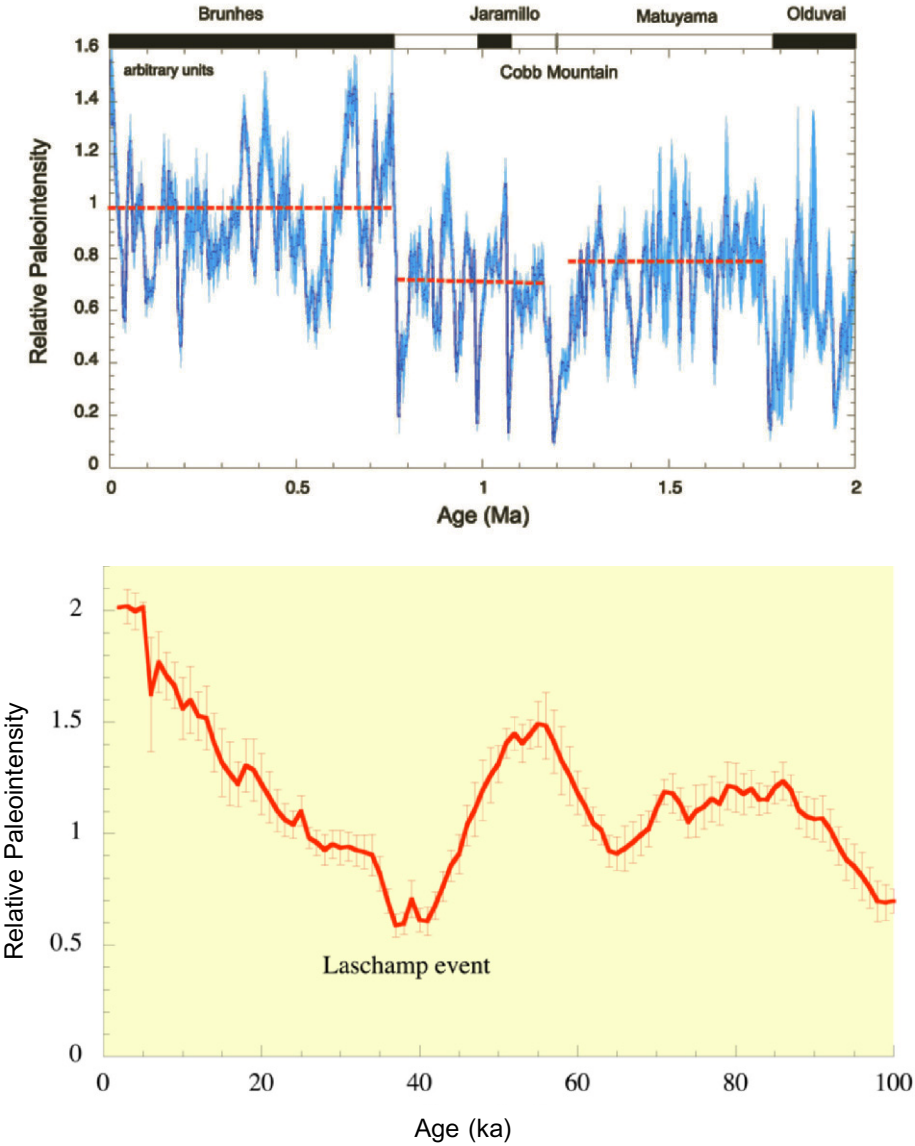


Figure 2.6 Top panel: Evolution of the intensity of the Earth magnetic field (in relative units) over the past 2 million years as a function of years before present. The horizontal unit is 100,000 years. The black and white bars at the top of the panel show the succession of the polarity intervals (black refers to the present or 'normal' polarity and white to the reverse). The names refer to either the pioneers of paleomagnetism (Brunhes, Matuyama) or to the sites where the measurements have been made. The horizontal red lines indicate the average intensity of the field for the various periods between reversals. The lower panel is an enlargement of the last 100,000 years. (*Credit*: J.P. Valet, reference [34].)

the magnetic field were to disappear completely, potentially severe effects might ensue. The solar wind might possibly erode and blow away a substantial amount of the Earth's atmosphere. However, this process might take a few million years, which is longer than the duration of the reversal itself and would not represent a genuine danger for life. Certainly, the magnetosphere would suffer important modifications with additional structural distortions from solar disturbances, creating at lower altitudes a radiation environment possibly harsher for life [35]. It has in fact been proposed that some life extinction events were associated with the disappearance of the Earth's magnetic field.

2.5 Evolution of the Earth's atmosphere

Tracing the evolution of the chemical composition of the early Earth's atmosphere during the Hadean and early Archean can be done through modeling of the process(es) that led to the formation of the planet and to isotope measurements in minerals. Contrary to the giant planets, because of its low gravity, the Earth, like Mercury, Venus and Mars, could not keep the two most abundant elements of the original solar nebula – hydrogen and helium – which were lost and returned to the interplanetary medium. Whatever remained of the other primitive gases has probably been swept out by the stronger solar wind of the young Sun.

During the Hadean, a secondary atmosphere was formed from the volatile compounds that were trapped into the accreting planetesimals that were outgassing from the molten rock [36]. Radioactive decay and impacts were the cause of the high temperatures of the early Earth. The outgassing of the original accreted material has been rapidly supplemented by ices and volatiles from the continuous impact–outgassing process, episodically delivering the basic elements of our atmosphere such as water-ice, carbon dioxide (CO_2), carbon monoxide (CO), methane (CH_4), ammonia (NH_3) and nitrogen [37]. The bombardment has also probably delivered a significant proportion of organic compounds, including amino acids which are also present in comets and meteorites.

At the same time as the Earth core formed from accretion, the metallic iron started to sink to the center allowing volcanic gases to release reduced species (devoid of oxygen) such as CH_4, NH_3 and H_2. The secondary atmosphere was certainly lost several times in the early history during the larger impacts, similar to the one that formed the Moon, and replenished through further outgassing and a less violent bombardment. The very hot surface temperature of the Earth did not allow water to remain liquid and the secondary atmosphere consisted mainly of water-steam together with these other gases. When the temperature dropped below 100°C, the water condensed and started to form the first oceans.

Molecular oxygen was almost absent in the early atmosphere, and consequently there was no ozone layer to prevent photodissociation of the secondary

atmosphere gases by the solar ultraviolet. Molecular oxygen and hydrogen atoms were, in this way, released through water photodissociation. The light hydrogen atoms escaped into interplanetary space, resulting in an increased atmospheric abundance of oxygen. If this were all that was happening, oxygen would accumulate indefinitely in the atmosphere. Fortunately, the oxidation of the gases, continuously replenished from continuing accretion together with volcanism, has consumed that oxygen. As a natural consequence of the planetary accretion processes of outgassing and oxidation, some kind of steady-state equilibrium could be reached, allowing the gradual build-up of an atmosphere containing N_2, CO_2, H_2O and O_2 in a proportion of about 0.005 the biogenic production (see Box 2.4).

Box 2.4 Where does the water come from?

The source of Earth's water remains unknown. Most probably, the planetesimals that formed the Earth were devoid of water because the solar nebula was too hot. Comets near and beyond the orbit of Jupiter (distances larger than 4 AU) are very numerous and by far potentially the most massive sources of water. It has been estimated that one million comets containing on average some 10^{15} kg of water would have been sufficient to create the first oceans. However, recent numerical simulations show that it is very unlikely that many of these objects could have collided with the Earth and that they rather scattered outward to the Oort Cloud and to the Kuiper Belt (Chapter 3). Furthermore, in three short-period comets (Halley, Hyakutake, Hale-Bop) the measured ratio of deuterium over hydrogen (D/H) has been found to be 3.1×10^{-4} or about twice that on Earth, adding doubts that the contribution of cometary water to the terrestrial oceans is larger than a few percent [38]. Carbonaceous chondritic material originating from the outer asteroid belt at distances larger than 2.5–3.5 AU may have delivered water to the Earth, but it is unlikely that they could have provided it all, based on the respective isotopic compositions of these bodies and of the Earth, unless an ad hoc existence of hydrated carbonaceous chondritic bodies is assumed with a composition closer to what we see on Earth, which is not impossible. An alternative possibility is that the Earth's water comes from much more massive impactors formed in the same asteroid belt as the carbonaceous meteorites which do present a D/H ratio closer to that of the Earth. The question is whether these giant collisions might not have freed more water than they supplied. Work is in progress to evaluate the plausibility of these various hypotheses.

If the Earth were devoid of an atmosphere, its effective surface temperature T_e, resulting from the thermal balance between heating from solar radiation at its present intensity and cooling from its own infrared emission to space, would be

254 K, well below the freezing point of water. The oldest zircons that must have formed while there was liquid water are dated at more than 4.3 billion years and indicate the presence of an ocean at that time. Hence, most of the time, it seems that the Earth has remained in the liquid-water regime. An icy Earth would have great difficulties to defrost and to create a liquid ocean because ice is a good reflector of solar light. Fortunately, the early atmosphere was able to create a greenhouse effect through which the infrared radiation emitted by the planet's surface, heated by solar radiation, is absorbed and re-emitted by infrared-active gases within the atmosphere (see Box 5.1: The *greenhouse effect*, on page 160)). This downward infrared radiation is able to warm the surface at a temperature of 14°C on average. In today's atmosphere, the most important greenhouse gases are CO_2 and water – the latter being responsible for nearly two-thirds of the effect. Other gases such as CH_4, NO_2, O_3 and the various anthropogenic chlorofluorocarbons contribute 2 to 3 degrees Celsius [37].

An early CO_2-rich atmosphere may have been present throughout the Hadean and the early Archaean [39], creating the strong greenhouse effect that allowed the temperature of the Earth to stay above the freezing point of water despite the fact that, according to the standard model of solar evolution [40], the early faint Sun was only about 70% as luminous as it is today. Figure 2.7 illustrates this 'Faint Sun' problem and the role of CO_2 in maintaining a global surface temperature T_s above the freezing point of water. About 0.3 bar of CO_2 would be needed to melt the ice of a totally frozen Earth [37]. However, as shown in Figure 2.7, it is likely that the atmosphere did not contain enough CO_2 to prevent the formation of an ice-covered ocean and form a 'snowball Earth'. If this has indeed occurred it is also likely that large impactors of ~100 km in diameter, which struck every 100,000–10 million years between about 3.6 and 4.5 billion years ago might have been strong enough to allow melting of an ice sheet of about 300 meters thickness, resulting in sets of thaw–freeze cycles associated with such impacts [36].

Another very powerful greenhouse gas is methane, 21 times as effective as CO_2, which might as well have contributed to the greenhouse effect. Methane is produced by biogenic and anthropogenic processes and can also be produced naturally by mid-ocean ridge volcanoes. There, water reacts with CO_2 releasing methane and molecular oxygen through the general equation:

$$CO_2 + 2H_2O \rightarrow CH_4 + 2O_2$$

However, the non-biogenic methane was probably very scarce in the Hadean era compared to its present concentration, and these local methane concentrations might not have been enough to substantially contribute to melting the snowball Earth. A feedback mechanism is also required to avoid an atmospheric 'runaway effect' that would result from the accumulation of CO_2 and water vapor in the atmosphere and would raise the temperature even further, accumulating more water and CO_2 and leading to a Venus-type situation (Chapter 9). This mechanism, called 'weathering', results from the property of CO_2 to dissolve in rain water to form the weak carbonic acid (H_2CO_3), strong enough however to

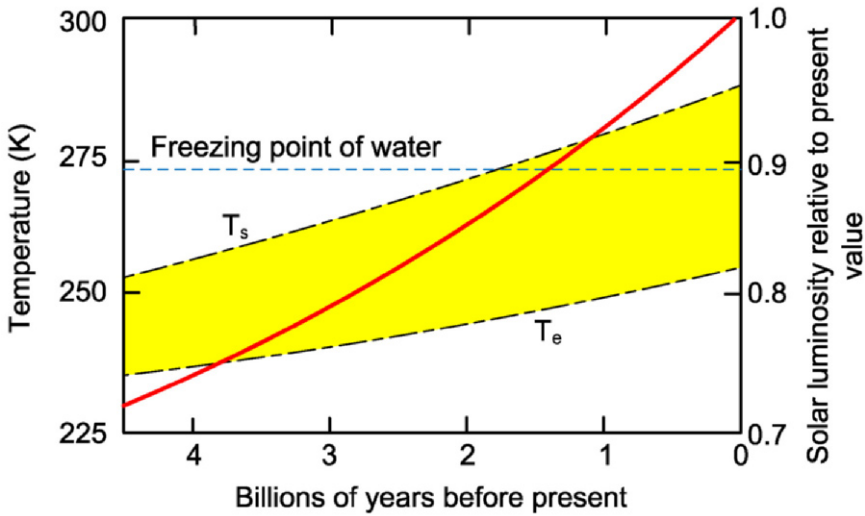


Figure 2.7 The faint young Sun problem. The red solid curve is a computed value of the solar luminosity relative to its present value. T_e is the Earth's effective radiating temperature without greenhouse effect presently equal to 254 K. T_s is the calculated mean global surface temperature assuming an amount of atmospheric CO_2 which has been fixed at 300 ppmv and a fixed relative humidity (see Kasting and Catling, reference [37].)

dissolve silicate rocks over long timescales, and create carbonates that accumulate in the ground through the general chemical equation:

$$CO_2 + CaSiO_3 \rightarrow CaCO_3 + SiO_2$$

The efficiency of this process is high and would eventually eliminate all CO_2 from the atmosphere, as was apparently the case on Mars, making the Earth, like the red planet, uninhabitable! Fortunately, plate tectonics played a determining role in keeping CO_2 in the atmosphere and maintaining an affordable greenhouse effect. The continuous creation and subduction of the seafloor at plate boundaries transports carbonate sediments to depths where the temperatures and the pressures are high and where they are destroyed through the inverse reaction that created them (called carbonate metamorphism), releasing new CO_2. The replenishment cycle of CO_2 in the atmosphere–ocean system through this process covers approximately half a million years, and is dependent on the surface temperature. At higher temperatures the evaporation of water increases while the precipitation and the concentration of CO_2 decrease. Conversely, it increases as the surface temperature falls. This set of complex mechanisms was essential in allowing the apparition and the development of life as we now describe.

2.6 Life and evolution

2.6.1 The early fossils in the Archean

Life appeared early on Earth, although 'how early' is a subject of much controversy. Abundant macroscopic fossils – the remains or imprints of organisms – provide the most direct evidence for past life and allow us to trace its evolution over the last 500–600 million years without too much ambiguity. Before that time the signs of life become sparser in a large part because the rocks containing the fossils have become much deformed and subjected to heat and volcanic processes. Nevertheless, in some places relatively unperturbed sediments are found in which earlier evidence for life has been preserved. There appears to be a consensus about the identification of fossils with ages of up to 2,000 million years and increasing, though far from unanimous, confidence in signs of life up to 3,500 million years. Interestingly this could bring the origin of life very close to the end of the Late Heavy Bombardment, which would have created conditions very hostile to life. The biological analyses of the most deeply branched organisms suggest that many early life forms were associated with hydrothermal systems that can either be of volcanic or of impact origin. The latter were probably more abundant than the former during the period of the bombardment. The volume of these systems can extend over the entire diameter of an impact crater and down to depths of several kilometers, providing suitable environmental conditions for thermophilic and hyperthermophilic forms of life. Either these were able to survive the conditions of the heavy bombardment, or the impact itself created the proper conditions for them to develop. Whether the very first life developed under high-temperature conditions or not has been extensively debated and has remained controversial.

One therefore gets the impression that, as soon as the Earth became liveable, life actually did appear. However, all this life was quite primitive: microbes, algae and organisms of uncertain parentage. It remained so until around the beginning of the Cambrian epoch ($\sim$540 million years BP) when, almost magically, representatives of most of the major divisions of life made their appearance. The long period that elapsed between the origin of life and the appearance of more advanced organisms remains somewhat mysterious. Probably early life had to remake the environment so that it became suitable for advanced life. As discussed previously, the early atmosphere was almost devoid of oxygen. It must have taken a long time for the appropriate microbes (cyanobacteria) to produce the required amount of oxygen. First it was used to oxidize the rocks in the Earth's crust and only later could the build-up of atmospheric oxygen begin. In fact, it is possible that it was only towards the Cambrian period that oxygen reached its current abundance in the atmosphere after a slow ascent, beginning in the earliest Proterozoic at about the time that the photosynthetically oxygen-producing cyanobacteria appeared.

The fossil record is incomplete. One only has to walk over a beach to see most of the dead jellyfish disappear in a few days, while most shells are ground up by the waves. Animals with shells or bony skeletons stand a better chance of being

preserved than animals with soft bodies. The latter are only preserved under special conditions of very rapid sedimentation and are only protected against rotting by being buried deeply enough. In some cases a landslide may have been responsible; in others a burial in lakes with bottom waters without oxygen. But even if successfully buried, later erosion may convert the whole layer of fossils into sand and clay without signs of the life that it contained at earlier times. As a result, while the presence of a certain class of fossil at a particular epoch shows that it existed at that time, its absence does not prove that it was not there.

These limitations may be partially overcome by two other types of evidence: isotope ratios in sediments and genetic comparisons of present-day organisms. In many biological processes different isotopes of elements behave slightly differently. For example, carbon compounds play an essential role in the construction of the cells of the organisms, and it turns out that biogenic carbon has a slight deficiency in the heavy isotope ^{13}C with respect to the (much more common) ^{12}C in inorganic matter. The reactions in which the different isotopes are engaged and the chemical compounds they lead to are the same, but the speed of the reactions may be just slightly different. So, if some carbon is found in ancient rocks, the $^{13}C/^{12}C$ ratio may document an organic origin. Similar effects occur in the sulfate cycles which are of much importance in various kinds of bacteria. The earliest such evidence for biological activity comes from 3,700-million-year-old minuscule graphitic globules that may represent the remains of planktonic organisms [41].

During the evolutionary process the genetic make-up of different species has gradually changed. When two species have a common ancestor, the genetic differences increase with the time that has elapsed since the split took place. But this common ancestor will also be the result of an earlier split, and by starting out with the species that are around today, one can construct a 'genetic tree' (Figure 2.8) in which the branches are chronologically arranged. Genetically very similar species generally have a recent common ancestor and are located on a common branch before the last split. Working backwards one comes to fewer branches deeper in time. Of course, it is not evident that the current speed of genetic change was the same in the past, but by the location of fossils on the various branches it is possible to calibrate the timescale for the branching points. The reality is more complex with different branches being possible, which makes it difficult to construct unique trees, but this should not significantly affect the timing of the earliest branch points.

Not surprisingly, the earliest fossils have been the subject of much controversy. Perhaps the most convincing are contained in 3,416 million-year-old rocks in South Africa which look like normal oceanic sedimentary deposits [42]. Both shallow and deeper water deposits have been found. Fine carbonaceous layers with filamentary structures look like microbial mats. The carbonaceous matter has $^{12}C/^{13}C$ ratios suggestive of a biological origin. The most interesting aspect is that these mat-like structures are only found in the shallow water deposits to a depth where sunlight can penetrate, suggesting that these were due to photosynthetic organisms, though not necessarily of an

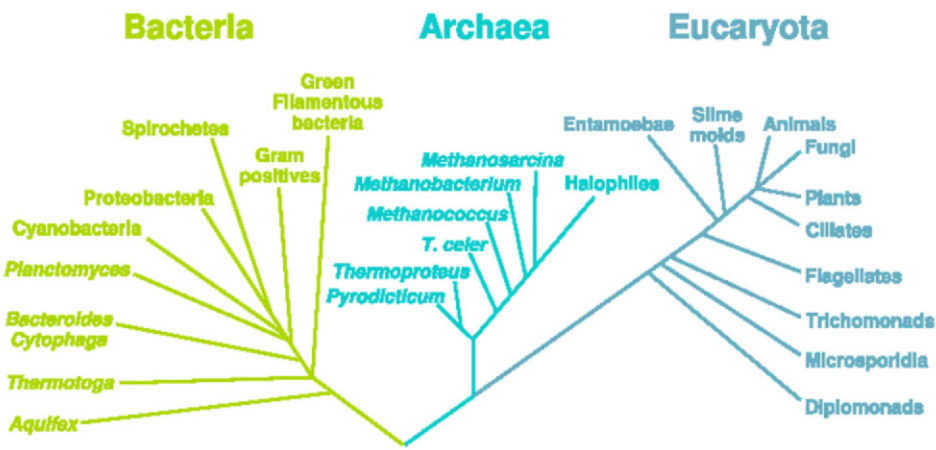


Figure 2.8 A Tree of Life. (*Credit*: Wikipedia.)

oxygen-producing kind. Around the same time, stromatolites appeared: layered structures which today are seen only in rather extreme environments. At present, stromatolites are biological structures composed of layers of algae and of sediment, but there has been uncertainty about the possibility that the 3.5-billion-year-old structures could be abiotic. However, recent studies support the view that they are, in fact, also of biological origin [43].

2.6.2 The Proterozoic and the apparition of oxygen

The next important step in the evolution of life was the development of oxygen-producing photosynthesis. Fossils of cyanobacteria are clearly present at 2,000 million years BP and possibly several hundred million years earlier. Clear biochemical evidence for their existence is found in 2,700-million-year-old Australian rocks [44].

Fossils of simple eukaryotic algae have been found in 2,100-million-year-old rocks near Lake Superior (Figure 2.9 [45]). But a more diverse and complex eukaryotic assembly appeared only 1,000–1,500 million years later. How can we explain such a long period of stasis after a promising beginning? An interesting suggestion is that a shortage of vital trace elements in the oceans may have held back evolutionary progress [46].

The earliest oceans presumably had much dissolved iron which was injected by hydrothermal sources. When this iron was deposited in sediments the characteristic 'banded iron' formations resulted, which still today are exploited to recover the metal. By 1,800 million years, these deposits had stopped, presumably because iron concentrations had become very low. Since the supply of iron must have continued, it apparently was transformed into an insoluble form. At first this was ascribed to the increase of oceanic oxygen which would have oxidized the iron, but now it is generally believed that most of the ocean

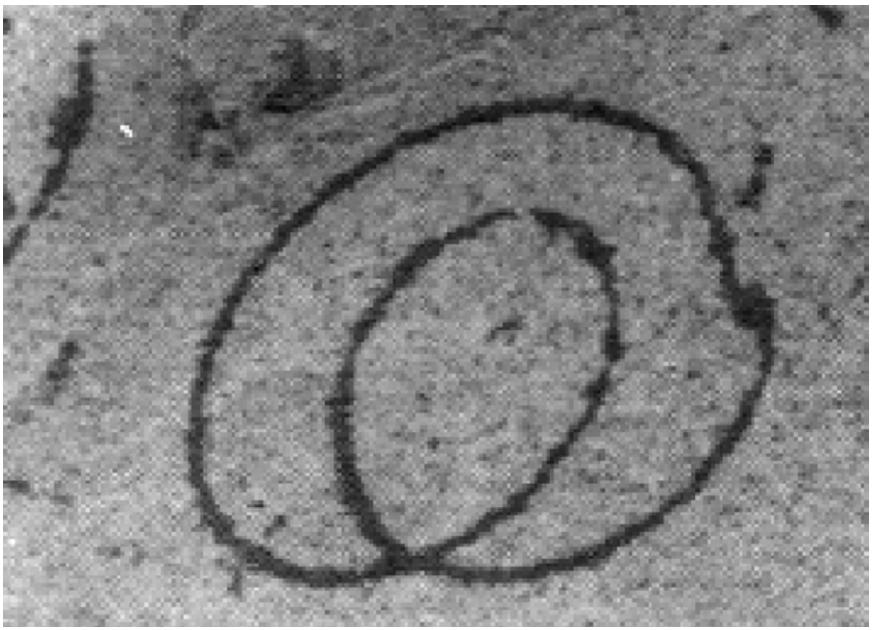


Figure 2.9 Specimen of Grypania from Neganaunee iron formation, Empire mine. (Han and Runnegar [45].)

was anoxic and had gradually become sulfidic [47]. As a result, the iron became incorporated in insoluble pyrite (FeS). Also, trace elements like molybdenum (Mo), copper and others were insoluble in the sulfidic oceans. These elements, and in particular molybdenum, play an essential role in living cells. Numerous enzymes contain the element including those that 'fix' nitrogen, converting atmospheric N_2 into ammonia. It has been suggested that the virtual absence of molybdenum may have constrained the further evolution of algae and eukaryotes in general, until the increase of the oxygen abundance led to the oxidation of the sulfides. This would have left molybdenum and some other trace elements with a present-day abundance level and allowed evolution to proceed. While this is still a speculative scenario, it illustrates the close connection between atmospheric, oceanic and biological evolution in which seemingly small causes may have large consequences, giving a hint that the Earth is a system. More specifically it shows the fundamental importance of the appearance of the cyanobacteria which still today maintain much of the world's oxygen.

2.6.3 The Neo-Proterozoic: the Ediacarans and the 'snowball earth'

About 575 million years ago abundant macroscopic fossils appeared, the Ediacarans, so named after some hills in Southeastern Australia, but later found to have a worldwide oceanic distribution. Without mineralized structures their preservation was only possible in special circumstances of very rapid burial under

Box 2.5 Atmospheric oxygen

The accumulation of oxygen in the atmosphere depends on the burial of organic carbon in sediments. Contrary to the popular belief in the forests as the 'lungs of the Earth', the contribution of land plants is negligible; the plants take CO_2 from the atmosphere and produce oxygen, but after their death they rot away consuming the oxygen and returning the CO_2 to the atmosphere. Only when their carbon-containing matter is locked up in sediments can there be a net gain in atmospheric oxygen, as happened during the deposition of the coal beds during the carboniferous period, when O_2 concentrations of 30% may have been reached. The full history of atmospheric oxygen is still rather uncertain and controversial. Its presence in the ocean has important consequences for the abundance of iron and of sulfides, and so studying ancient oceanic chemistry inferences about past oxygen concentrations may be made.

Probably, concentrations remained very low until the 'Great Oxidation Event' about 2,400 million years ago, when values at least 1% to 10% of present were attained. This is 300 million years or more after oxygen-producing cyanobacteria originated. By the time that the first Ediacarans appeared concentrations had climbed to more than 15% of present values. It is likely that so much oxygen was needed for the functioning of sizable animals. Over the last 500 million years oxygen levels remained within a factor of 2 from present-day values.

The stability of O_2 in the atmosphere may be seen as follows. Suppose that suddenly all forests and land plants were burned. This would add some 300 ppm of CO_2 to the present 370 ppm. Since in the process every carbon atom would have combined with an O_2 molecule, the O_2 concentration (in mole fraction) would have diminished by 370 ppm. But the O_2 concentration is 21% or 209,000 ppm which therefore would change less than 0.2%. Only over geological timescales are large changes in O_2 concentration likely to occur.

sediments. They represent a rather bizarre fauna (Figure 2.10), mostly without obvious connections to later types. The typical specimens are fern-like fronds of a few centimeters up to more than 1 meter in size. In some cases they look like having a pedestal suggesting that they fixed themselves to the sea bottom. Slightly younger groups include centimeter-sized round disks and disks with trilateral symmetry [48].

The nature of most Ediacarans is still controversial. Some have argued that they may represent a failed attempt in the evolution of more complex organisms without connection to present-day animals, but others have concluded that connections to later phyla may be made [49]. In any case, the typical Ediacaran



Figure 2.10 Some Ediacaran fossils. (a) Dickinsonia, seen by some as related to the worms or the corals; (b) Spriggina, a possible arthropode; (c) Charnia, a frond-like organism. Typical sizes are several centimeters or somewhat more. The association with present-day animal phyla, if any, is still very uncertain. (*Credit*: Wikipedia.)

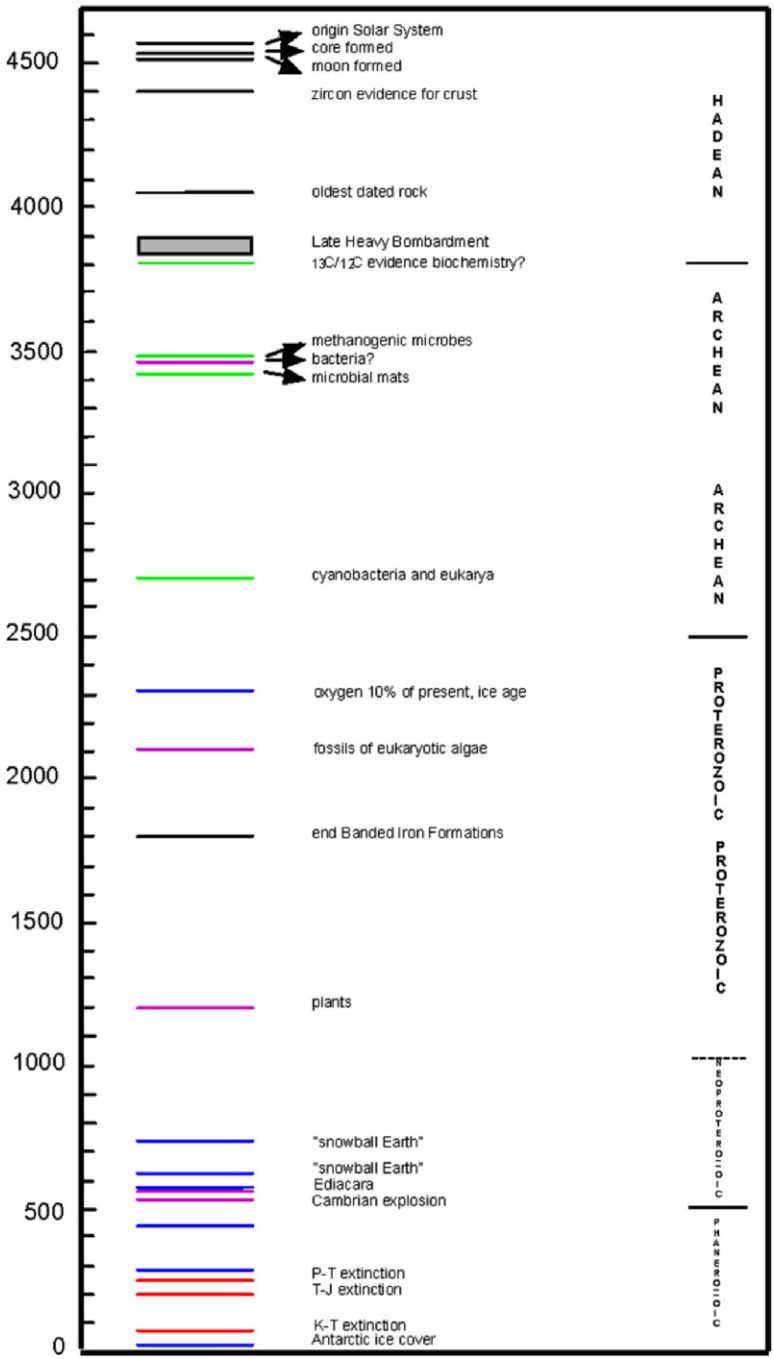


Figure 2.11 A summary of events during the Earth's history.

had largely disappeared when the 'Cambrian explosion' in animal diversification arrived. At the beginning of the Cambrian geological period within perhaps no more than 20 million years, many of the major divisions (phyla) appeared in the fossil record. Most of these showed no obvious connection with the preceding Ediacarans. Was an environmental cause responsible for this sudden blossoming of more advanced life?

Ediacaran times are characterized by major climatological and biochemical events for which some dates have recently become available. Just preceding the Ediacaran was the so-called Marinoan glaciation. Paleomagnetic reconstructions indicate that the glacial deposits reached close to the equator. The end of the glaciation is defined by the deposition of a layer of carbonates ('cap carbonates') which has now been dated at 635 million years and marks the beginning of the Ediacaran. A last glaciation occurred 580 million years ago. The typical Ediacarans would have appeared just after the last glacial. The terminations of these two coincided with excursions in the $^{13}C/^{12}C$ record that have been taken to indicate much reduced biological activity. This has been interpreted as evidence for 'Snowball Earth', or total glaciation of the whole Earth [50]. Under the ice, only rudimentary life would be possible, but if the Earth were frozen over, no exchanges of CO_2 with the oceans would take place, while volcanoes would continue to inject the gas into the atmosphere. Volcanic CO_2 would then have built up above the ice to very high concentrations (100,000 ppm) with the resulting greenhouse effect leading to a melting of the ice. Once the ice began to melt the reflectivity of the Earth would diminish and, as a result, the melting would accelerate. The high concentration of CO_2 would continue to lead to a strong greenhouse, and temperatures would be high after the ice was gone. The ample CO_2 then led to the deposition of the carbonates, until the CO_2 had been returned to more typical concentrations. While it is generally agreed that these ice ages, in particular the Marinoan and at least one before, were very severe and that much of the continental surface was covered with ice, it is still unclear whether the oceans were covered as well. In any case, the relation of the evolution of the Ediacarans and the earliest metazoans with the climatological events is still obscure.

The rise of oxygen in the atmosphere and ocean must have had a major impact on the evolution of more complex life. It is generally concluded by biologists that larger animals could only have come into existence on the basis of oxygenic respiration. Certainly oxygen abundances increased towards the beginning of the Cambrian, but present data seem inadequate to make a more precise correlation with the detailed evolution of life.

An early mineralized tube-forming animal, Cloudina, appears in remarkably well-preserved late Ediacaran rocks in southern China. Of particular interest are the small holes bored in several of the tubes which suggest that predators were already at work [51]. Also, late in the Ediacaran small shelly animals become more abundant. So, predation may have been an important factor pushing organisms towards stronger mineralized protective shields, which at the same time greatly increased the chance of being fossilized. Could predation also have

contributed to the extinction of the rather unprotected and not very mobile frond-like Ediacaran fauna? Some 10–20 million years after the end of the Ediacaran, life proliferated and many of the phyla appeared which, today, dominate the macroscopic biological world. Even relatively advanced fish-like animals appeared that looked a bit like lampreys from which, much later, amphibians, reptiles and mammals developed [52].

The reasons why life rather suddenly proliferated have been extensively debated, with genetic mechanisms also being considered more recently [53]. The essential idea would be that the genetic machinery had first to evolve to an adequate state to be able to make different combinations corresponding to different animal types. Once it had done so, the different groups developed rather quickly and ecological interaction (predation, etc.) led to optimized fitness depending upon a limited number of variables. Once these had all been explored by the rearrangements of genetic elements, there was little possibility for drastically new body plans and so, even after great extinctions, no fundamentally new phyla appeared.

In the subsequent 500 million years, evolution continued, ultimately producing the biological world we know today. Along the way, nearly as many species became extinct as appeared, but no radically new body plans appear to have developed. Different species became abundant at different times and the locations at which they were first found have given the name to some of the geological epochs for which they were characteristic. Thus, the geological timescale was essentially biological. However, in the meantime radioactive dating has provided the absolute ages. Since the epochs were defined by characteristic species, it is no surprise that the separations between them correspond to extinctions.

2.6.4 The Phanerozoic, life extinctions

New species came about as a result of genetic mutations which were favorable in view of ecological opportunities. Extinctions followed when newer species took over or when environmental conditions deteriorated for a particular species. But from time to time bigger events happened: many species became extinct simultaneously and thereafter a variety of new species found the ecological space to appear in short order. Four major extinctions occurred in the Ordovician (420 million years), at the end of the Permian (250 million years), at the end of the Cretaceous (65 million years), and somewhat less catastrophic at the end of the Triassic (200 million years) [54]. Numerous smaller events have been identified or claimed by various authors.

Averaged over tens of million of years, the rates of extinction and of formation of new species have been more or less in balance, except in the very recent past when human influences have caused a large increase in the extinction rate – an increase of no less than a factor of 100–1,000! In the end of Cretaceous extinction, significant sections of the biological world disappeared. The coiled ammonites, so abundant in marine communities during the Mesozoic, and the dinosaurs which had dominated the land, all disappeared never to be seen again.

As much as half the number of all species did vanish in a very short time. The end of the Permian extinction was perhaps even more drastic. However, while the end of the Cretaceous extinction seems to have been virtually instantaneous, the end of the Permian one appears to have been more complex extending over several million years [55]. Some 70% of species may have been wiped out. A wide variety of causes has been invoked, ranging from a nearby supernova (see Chapter 3) to climate variations. However, three possibilities have been more extensively discussed in recent times: impacts of asteroids or comets, flood basalts and anoxic conditions.

There is little doubt that at the end of the Cretaceous 65 million years ago (K–T extinction in Figure 2.11), a substantial extraterrestrial body hit the Earth. The boundary to the Tertiary is defined by a thin centimeter scale layer of clay in which the element iridium is a factor of 250 more abundant than normal in the Earth's crust [56]. A very large overabundance of iridium is found in meteorites and should also apply in asteroids. It has been generally thought that the lower abundance in the crust is due to the iridium having settled in the Earth's iron–nickel core. If an asteroid hit the Earth, it would have vaporized and much of its iridium-rich material would have rained down somewhat later. Similar observations were made over nearly 100 sites on Earth. The amount of iridium deposited in that worldwide layer corresponds to that expected in an asteroid of 10–14 km diameter. Indeed, a 65-million-year-old crater of some 180 km diameter was found in 1991 by a Mexican oil-prospecting company at the bottom of the Gulf of Mexico in the area near Chicxulub, a fishing village on the north coast of Yucatan [57]. The size of the crater corresponds to the impact of such a massive asteroid.

When a massive asteroid hits, a huge crater is excavated. The collision of a 10-km object with the Earth would raise the temperature to several thousand degrees and the atmospheric pressure a million times. This enormous energy, equivalent to several tens of millions of megatons of TNT, fractures the rocks and vitrifies the ground, creating small spheres of glass and small crystals, and highly shocked quartz. Quartz is a very stable crystal and only huge energy dissipation can alter its structure. The presence of fractures in quartz crystals can only be explained by shock waves of a tremendous energy such as those involved in the bombardment of an asteroid. In addition, high concentrations of soot and ashes, several hundred times higher than the average, resulting from the gigantic fires that inevitably occur after such cataclysms, are sent into the atmosphere. Such signatures have been found which correspond to the end of the Cretaceous.

That was not a pleasant moment for the dinosaurs. Those that were just near the impact zone may have heard the tremendous noise of the shock wave as the object was speeding more than 70,000 km/h towards the tropics. The impact resulted in a gigantic tsunami, whose height has been evaluated at 100–300 meters, accelerated outward through the ocean, thousands of kilometers away, drowning all living species on the littorals and ripping up seafloor sediments down to depths of 500 meters. The air blast flattened any forests within a 1,000–2,000-km-diameter region. Computer modeling indicates that the initial impact

generated an earthquake of 13th magnitude on the Richter scale, 3,000 times stronger than the largest known, which pervaded the entire Earth [58]. The atmosphere was filled with millions of tons of debris. Within a few hundred kilometers they accumulated in a layer several hundreds of meters thick, enough to totally cover and exterminate any local life. The most significant perturbations probably came from a gigantic plume of vapor and high-velocity debris that arose above the atmosphere, with velocities large enough to put them in orbits around the Earth. Some of this material subsequently re-accreted to the Earth and later settled down to the ground as the result of atmospheric friction. When they impacted the ground, they ignited intense fires. The effects on forests would depend in part on the season. This second burning would kill what remained of the living species. The consequences, however, were more global [59]. The sky was obscured as the light from the Sun could no longer reach the ground and the Earth plunged into an artificial winter that lasted several months or years [60], although the oceans might have been less affected because of their large thermal inertia. As photosynthesis stopped for approximately a year, with a consequent disruption of the marine food chain, massive death most probably followed, and the temperature became polar everywhere.

In addition, several greenhouse gases, including carbon dioxide, sulfur and chlorine that were trapped in the shocked rocks went into the atmosphere filling it to an amount five orders of magnitude greater than what would be necessary to destroy the ozone layer. Based on measurements of soot deposited in sediments, it has been estimated that the fires released some 10,000 Gt* of CO_2, 100 Gt of methane and 1000 Gt of carbon monoxide, which correspond to 300 times the amount of carbon produced annually by the burning of fossil fuels at present.

Clearly, both the primary and the secondary effects of the impact left very little chance of life. The most resistant species might have survived the first shock, but they probably could not withstand the infernal warming that followed, as clouds of greenhouse gases built up. Re-establishing food chains and integrated ecosystems making the Earth liveable again would take several decades or centuries. In addition to the largest population on the planet at that time – the dinosaurs – other species also disappeared: pterosaurs, plesiosaurs, mesosaurs, ammonites and many categories of fish and plants. No more than 30% of the living species could apparently survive. The flux of organic material to the deep sea took about 3 million years to recover [61]. The consequence of all of this would have been a large quantity of dead organic matter, and this appears to be confirmed by a brief proliferation of fungi at the boundary layer [62].

An intriguing scenario for the origin of the end of Cretaceous impactor has been developed. It begins with an asteroid family with, as its largest member, the 40-km-diameter Baptistina [63]. An analysis of the motion of the members of the family suggests that they are the result of a collision in the asteroid belt between

* One gigaton or Gt is equal to 10^9 tons.

Mars and Jupiter some 160 million years ago and that several fragments have been placed on Earth-crossing orbits. The similarity of the composition of Baptistina to that inferred for the impactor makes it tempting to identify the K–T object as a relic of that collision. As the authors of a comment on this story wrote [64], 'It is a poignant thought that the Baptistina collision some 160 million years ago sealed the fate of the late-Cretaceous dinosaurs, well before most of them had even evolved.' Are there still more remnants of the collision lurking on Earth-crossing orbits?

Curiously, at about the time of the impact a huge eruption of basaltic magma occurred in India, producing the 'Deccan traps', which represent an outflow of some 3,000,000 km^3 of magma over a time of the order of a million years [65]. Such outflows pour out of large cracks in the Earth's crust. During historical times this type of volcanism has occurred in Iceland – the Laki eruption of 1783 – which, by its poisonous vapors killed directly or indirectly a quarter to half of the Icelandic population [66]. But Laki produced just 12 km^3. Recent data seem to indicate that the high point of the Deccan eruption occurred several 100,000 years before the asteroid struck [67]. So it would seem that the latter is the prime culprit, although the lava vapors may have already placed the biosphere under stress before it hit.

After the end-Cretaceous extinction had been ascribed to the impact event, many papers were written which claimed to connect major and minor extinctions with more or less contemporaneous impact craters. However, with the uncertain dates and the high frequency of impacts, much doubt has remained. In particular, the severe end-of-Permian crisis (P–T extinction in Figure 2.11) has not convincingly been connected to one or more impacts. However, a major flood basalt eruption at least as large as the Deccan traps – the Siberian traps – occurred irresolvably close in time [68]. In addition, there is evidence that the oceans at the time were anoxic, which could have resulted from the fact that most continents were largely united in one great block that would have weakened the vertical oceanic circulation and substantially reduced the available habitats.

On a much smaller scale, the dangers of an anoxic layer at the bottom of a body of water have become apparent in Lake Nyos in Africa [69]. The pressure of the gas made the deep CO_2-rich layer unstable, and led to a large cloud of CO_2 escaping. Since CO_2 is heavier than air, the CO_2-rich layer stayed relatively close to the ground and asphyxiated thousands of people. Is it possible that a massive escape of CO_2 (and also H_2S) from the deep ocean could have caused the (multiple?) extinctions at the end of the Permian? The matter might also have been greatly aggravated by the CO_2 and sulfide injections from the Siberian traps. Also the active injection of sub-oceanic magma or the impact of an asteroid could set off the overturning process. The different mechanisms, therefore, are not necessarily mutually exclusive.

A third very large flood basalt – the Central Atlantic Magmatic Province (CAMP) – covers parts of the Americas, Africa and Europe. CAMP originated at the time the Atlantic Ocean began to open up in the Gondwana super-continent [70]

(Figure 2.4). Perhaps even bigger than the Deccan flows, CAMP was deposited around 200 million years ago, at the time of the end of Triassic mass extinction events. Thus, the three largest mass extinctions of the last 400 million years seem all to be associated with the three largest flood basalt eruptions of the last 500 million years. Coincidence?

In the quantitative analysis of the extinctions there are serious problems associated with the incompleteness of the fossil record. Relatively rare species are more likely to be classified as extinct than more abundant ones, since the chances of preservation are so small. Even more important may be the availability of sedimentary outcrops where one can search for fossils. For example, if the sea level declines the coastal sediments from the preceding period may be eroded away, and so no fossils of coastal organisms of that period would be found, even though corresponding species had survived. Some of the extinctions may therefore be weaker than the records suggest. Of course, such effects cannot explain why, after the end of the Cretaceous, not a single dinosaur or ammonite was ever seen again. So some mass extinctions are undoubtedly real, but the sharpness of the extinction spike and its quantitative importance may have been affected by the incompleteness of the records.

Although the mass extinctions were a disaster for some species, they created the ecological space in which others could flourish. The development of the dinosaurs might have been favored by the huge event at the end of the Triassic which freed many ecological niches previously occupied by other species, allowing the dinosaurs to flourish in the wake of that massive and sudden extinction. And the extinction of the dinosaurs themselves possibly created the opportunity for the mammals to proliferate. Small mammals had already existed for perhaps 100 million years before the dinosaur extinction, so whether they would have come to prominence in any case will remain an open question.

At present a new mass extinction appears to be on its way, although it is qualitatively different from the preceding ones. Then the extinction was caused by events outside the biosphere which created ecological space for new developments. This time it is the sudden proliferation of one species – homo – at the expense of the rest of the biosphere, partly by predation and partly by removing ecological space. The beginning of the current extinction wave is visible in Australia where several species of animals became extinct 60,000 years ago when humans arrived. It has been a subject of controversy whether humans were responsible for climate change, but recent research has tended to confirm the former [71]. At the end of the last ice age many species of large animals who had happily survived the past glacial endings, became extinct at the time that humans became more practiced in the use of tools. Again, much controversy has surrounded the subject, but the evidence, in the Americas in particular, seems to go in the direction of human predation. And, of course, today the evidence is there for all to see with the oceans being emptied of many species of fish and whales, and the continents of amphibians, birds and mammals. While it is true that many species are 'on the brink of extinction' and could still recover if drastic

measures were taken, in many cases the small remaining populations may be too small to resist the unavoidable environmental stresses, sickness and other factors that endanger small populations. In the case of humans, disease also played a role in the extinctions of small populations, but whether it was a factor in past extinctions in the biosphere in general is unknown. However, the general species' specific pathogen character of viruses and bacteria has been thought to make it improbable that disease could be a major factor in mass extinctions. The recovery of some Amerindian populations, after having been decimated by European diseases, also suggests that disease plays a minor role in the extinctions of larger populations. This topic will be again discussed in Chapter 4.

2.7 Conclusion

The future of many species on Earth will be decided in the current century. It is contingent on two factors: (1) to leave enough space for the wild fauna and flora and (2) to put an end to uncontrolled hunting, logging and collecting. The tendency of humans to modify nature to accommodate increasing numbers of their own kind, and the rewards of hunting and logging, are such that only in a well-regulated world society is there much hope to avoid the extinction of the larger species and of many smaller ones as well. A peaking of the world population in the future could help to create the necessary space, but in the meantime what has already gone is lost forever. So, returning to our 100,000 year world, much of its 'nature' will depend on what is done during this century. Would it not be worth while to try *not* to leave a totally depleted biosphere? But before addressing this most essential political issue, we must look at what might occur to our planet under the threats originating in the cosmos and those that our planet creates itself.

2.8 Notes and references

[1] Lunine, J.I., 1999, *Earth, Evolution of a Habitable World,* Cambridge University Press, p. 319.
[2] Wilde, S.A. et al., 2001, 'Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago', *Nature* **409,** 175–178; see also Valley, J.W., 2005, 'A cool early Earth?', *Scientific American* **293** (4), 40–47.
[3] Zimmer, C., 1999, 'Ancient continent opens window on early Earth', *Science* **286**, 2254–2256.
[4] Jacobsen, S.B., 2003, 'How old is planet Earth?', *Science* **300**, 1513–1514.
[5] Fitzgerald, R., 2003, 'Isotope ratio measurements firm up knowledge of Earth's formation', *Physics Today* **56** (1), 16–18.
[6] Watson, E.B. and Harrison, T.M., 2005, 'Zircon thermometre reveals minimum melting conditions on earliest Earth', *Science* **308**, 841–844.

[7] Nisbet, E.G. and Sleep, N.H., 2001, 'The habitat and nature of early life', *Nature* **409**, 1083–1091.
[8] The size of craters depends not only on the characteristic dimensions of the impactor but also on the gravity of the impacted body: because gravity acts as a focusing attraction force, the smaller the gravity field the larger the diameter of the crater. This explains why the Moon possesses larger craters than the Earth.
[9] Samples of the lunar surface have revealed three major types of rock-chemistry: anorthositic crust, mare basalts and the KREEP (acronym for Potassium-Rare Earth Elements Phosphorus elements that are enriched by a Moon-specific differentiation process). The chemistry of these rock types has allowed us to reconstruct the differentiation processes and establish the chemical composition of the total Moon. Many laboratories and institutes were involved in lunar rocks studies. The absolute timescale of lunar events was established in the USA (e.g. Caltech, UCSD, Stony Brook) and in Europe (University of Bern, Sheffield, Paris and MPI Heidelberg). In addition, heat flow and seismic measurements, gamma-ray and x-ray surveys by the Apollo lunar orbiter and orbital information as influenced by gravity field anomalies provides supplementary geophysical information (Geiss, J., private communication).
[10] Lodders, K. and Fegley, B., 1998, *The Planetary Scientists Companion,* Oxford University Press, p. 382.
[11] Taylor, S.R., 1982, *Planetary Science: A Lunar Perspective*, Lunar and Planetary Institute, Houston, Texas, p. 481.
[12] Anderson, D.L., 1989, *Theory of the Earth*, Blackwell Publications, Boston, p. 366.
[13] Canup, R.M. and Righter K. (eds) 2000, *Origin of the Earth and Moon,* University of Arizona Press, Tucson, p. 555.
[14] Geiss, J., 2000, 'Earth Moon and Mars', *Spatium* **5**, Association Pro-ISSI Pub., 3–15
[15] Wills, C. and Bada, J., 2000, *The Spark of Life: Darwin and the Primeval Soup*, Perseus, p. 330.
[16] Laskar, J. et al., 1993, 'Stabilization of the Earth's obliquity by the Moon', *Nature* **361**, 615–617.
[17] Morbidelli, A. et al., 2005, 'Chaotic capture of Jupiter's Trojan asteroids in the early Solar System', *Nature* **435**, 462–465.
[18] Hartman, W.K. et al., 2005, 'Chronology and physical evolution of planet mars', in *The Solar System and Beyond: Ten Years of ISSI*, ISSI book series, 211–228.
[19] McCauley, J.F. et al., 1981, 'Stratigraphy of the Caloris Basin, Mercury', *Icarus* **47**, 184–202.
[20] Schoenberg, R. et al., 2002, 'Tungsten isotope evidence from ~3.8-Gyr metamorphosed sediments for early meteorite bombardment of the Earth', *Nature* **418,** 403–405.

[21] Gomes, R. et al., 2005, 'Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets', *Nature* **435,** 466–469.
[22] Koeberl, C., 2006, 'The record of impact processes on the early Earth: A review of the first 2.5 billion years', *Geological Society of America*, *Special Paper* **405,** 1–23
[23] Campbell, I.H. and Taylor, S.R., 1983, 'No water, no granites, no continents', *Geophysical Research Letters* **10**, 1061–1064.
[24] Ashwal, L.D., 1989, Growth of continental crust: An introduction. *Tectonophysics* **161**, 143–145. (Courtesy, C. Heubeck.)
[25] Taylor, S.R. and McLennan, S.M., 1995, 'The geochemical evolution of the continental crust', *Reviews of Geophysics* **33**, 241–265.
[26] Wegener, A., 1924, *The Origin of Continents and Oceans*, Methuen, London, p. 276.
[27] Torsvik, T.H., 2003, 'The Rodinia Jigsaw puzzle', *Science* **300**, 1379–1381.
[28] Muir, H., 2003, 'Hell on Earth', *New Scientist* **2424**, 36–37.
[29] By courtesy of G.A. Glatzmaier, Earth and Planetary Sciences Department, University of California Santa Cruz, CA 95064 USA.
[30] Valet, J.P. and Courtillot, V., 1992, 'Les inversions du champ magnétique terrestre', *La Recherche* **246,** Vol. 23, 1002–1013.
[31] Dormy, E., 2006, 'The origin of the Earth's magnetic field: fundamental or environmental research?', *Europhysics News* **37–2** , 22–25.
[32] Monchaux, R. et al., 2007, 'Generation of a magnetic field by dynamo action in a turbulent flow of liquid sodium', *Physical Review Letters* **98,** 044502(4).
[33] Berhanu, M. et al., 2007, Magnetic field reversals in an experimental turbulent dynamo, *Europhysics Letters* **77**, 59001(5).
[34] Valet, J.P. et al., 2005, 'Geomagnetic dipole strength and reversal rate over the past two million years', *Nature* **435**, 802–805.
[35] Vogt, J. et al., 2004, 'MHD simulations of quadrupolar magnetospheres', *Journal Geophys. Res*. **109**, Issue A12, A12221, p. 14.
[36] Botta, O. and Bada, J.L., 2002, 'Extraterrestrial organic compounds in meteorites', *Surveys in Geophysics* **23**, 411–467.
[37] Kasting, J.F. and Catling, D., 2003, 'Evolution of a habitable planet', *Annual Review of Astronomy and Astrophysics* **41,** 429–463.
[38] Robert, F., 2001, 'The origin of water on Earth', *Nature* **293,** 1056–1058.
[39] Kasting, J.F., 1993, 'Earth's early atmosphere', *Science* **259,** 920–926.
[40] Gough, D.O., 1981, 'Solar interior structure and luminosity variations', *Solar Physics* **74,** 21–34.
[41] Rosing, M.T., 1999, '13C-depleted carbon micro particles in >3700-Ma sea-floor sedimentary rocks from western Greenland', *Science* **283**, 674–676.
[42] Tice, M.M. and Lowe, D.R., 2004, 'Photosynthetic microbial mats in the 3,416 million years-old ocean', *Nature* **431**, 549–552. Also: Westall, F. et al., 2006, 'Implications of a 3.472–3.333-Gyr-old subaerial microbial mat from the Barberton greenstone belt, South Africa for the UV environmental

conditions on the early Earth', *Philosophical Transactions of the Royal Society* **B 361**, 1857–1875.
[43] Allwood, A.C., 2006, 'Stromatolite reef from the early Archaean era of Australia', *Nature* **441**, 714–718.
[44] Brocks J.J. et al., 1999, 'Archean molecular fossils and the early rise of Eukaryotes', *Science* **285**, 1033–1036, and the commentary in Knoll, A.H., 1999, 'A new molecular window on early life', *Science* **285**, 1025–1026.
[45] Han, T.M. and Runnegar B., 1992, 'Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee iron-formation, Michigan', *Science* **257**, 232–235.
[46] Anbar, A.D. and Knoll, A.H., 2002, 'Proterozoic ocean chemistry and evolution: a bioinorganic bridge?', *Science* **297**, 1137–1141.
[47] Poulton, S.W. et al., 2004, 'The transition to a sulphidic ocean $\sim$ 1.84 billion years ago', *Nature* **431**, 173–177.
[48] Cloud, P. and Glaessner, M.F., 1982, 'The Ediacaran period and system: Metazoa inherit the Earth', *Science* **218**, 783–792.
[49] Conway Morris, S., 1993, 'The fossil record and the early evolution of the metazoa', *Nature* **361**, 219–225.
[50] Hoffman, P.F. et al., 1998, 'A neoproterozoic snowball Earth', *Science* **281**, 1342-1346; also: Donnadieu, Y. et al., 2004, 'A 'snowball Earth' climate triggered by continental break up through changes in runoff', *Nature* **420**, 303–306.
[51] Bengtson, S. and Zhao, Y., 1992, 'Predatorial borings in late Precambrian mineralized exoskeletons', *Science* **257**, 367–370.
[52] Shu, D.-G. et al., 1999, 'Lower Cambrian vertebrates from south China', *Nature* **402**, 42–46; Chen, J.-Y. et al., 1999, 'An early Cambrian craniate-like chordate', *Nature* **402**, 518–522.
[53] Marshall, C.R., 2006, 'Explaining the Cambrian 'explosion' of animals', *Annual Review of Earth and Planetary Sciences* **34**, 355–384.
[54] Sepkoski, J.J., 1995, *Global Events and Event Stratigraphy* (Ed. O.H. Walliser), Springer Verlag, Publ., pp. 35–57.
[55] Erwin, D.H., 1994, 'The Permo-Triassic extinction', *Nature* **367**, 231–236.
[56] Alvarez, W. et al., 1990, 'Iridium profile for 10 million years across the Cretaceous-Tertiary boundary at Gubbio (Italy)', *Science* **250**, 1700–1701.
[57] Morgan, J. et al., 1997, 'Size and morphology of the Chicxulub impact crater', *Nature* **390**, 472–476.
[58] Busby, C. et al., 2002 'Coastal landsliding and catastrophic sedimentation triggered by Cretaceous–Tertiary bolid impact: a Pacific margin example?', *Geology,* Geological Society of America, **30**, (8), 687–690.
[59] Kring, D.A., 2000, 'Impact events and their effects on the origin, evolution and distribution of life', *GSA Today*, Geological Society of America Publ. **10** (8), 1–7.
[60] Pollack, J.P. et al., 1983, 'Environmental effects of an impact-generated dust cloud: implications for the Cretaceous-Tertiary environmental effects of an impact generated dust cloud: Implications for the Cretaceous–Tertiary extinctions', *Science* **219**, 287–289.

[61] Kring, D.A. and Durda, D.D., 2001, 'The distribution of wildfires ignited by high-energy ejecta from the Chicxulub impact event', *Lunar and Planetary Science* **XXXII,** 1–2.
[62] Vajda, V. and McLoughlin, S., 2004, 'Fungal proliferation at the Cretaceous-Tertiary boundary', *Science* **303**, 1489.
[63] Bottke, W.F. et al., 2007, 'An asteroid break up 160 million years ago as the probable source of the K/T impactor', *Nature* **449**, 48–53.
[64] Claeys, P. and Goderis, S., 2007, 'Lethal billiards', *Nature* **449,** 30–31.
[65] Courtillot, V.E. and Renne, P.R., 2003, 'On the ages of flood basalt events', *Compte Rendus Géoscience* **335** (1), 113–140.
[66] Stone, R., 2004, 'Iceland's doomsday scenario?', *Science* **306**, 1278–1281.
[67] Ravizza, G. and Peucker-Ehrenbrink, B., 2003, 'Chemostratigraphic evidence of Deccan volcanism from the marine osmium isotope record', *Science* **302**, 1392-1395.
[68] Reichow, M.K. et al., 2002, '$^{40}Ar/^{39}Ar$ dates from the West Siberian basin: Siberian flood basalt province doubled', *Science* **296,** 1846–1849.
[69] Freeth, S.J. and Kay, R.L.F., 1987, 'The lake Nyos gas disaster', *Nature* **325**, 104–105.
[70] Marzoli, A. et al., 1999, 'Extensive 200-million-year-old continental flood basalts of the Central Atlantic Magmatic Province', *Science* **284**, 616–618.
[71] Barnosky, A.D. et al., 2004, 'Assessing the causes of late Pleistocene extinctions on the continents', *Science* **306**, 70–75.

3 Cosmic Menaces

These terrors, this darkness of the mind, do not need the spokes of the Sun to disperse, nor the arrows of morning light, but only the rational study of nature.

Lucretius

3.1 Introduction

Since 3.5 billion years ago, life has developed to a high level of sophistication on a planet sitting at the right distance to its star, orbited by a fortuitous moon which formed very early in the planet's history, the result of a gigantic collision which helped to stabilize the climate and the setting of conditions that give support to human development. Thereafter, life has evolved as a combination of Darwinian adaptation and of natural traumas which led to the various extinctions mentioned in the previous chapter. Will such events recur in the future or, more specifically, in the next 100,000 years that may ruin all possible efforts of maintaining the Earth in a habitable state for humans? What might these events be? If we know what they are can we protect ourselves from their occurrence and avoid their disastrous effects? Besides those that are anthropogenically generated, there are two main types of natural hazards. In this chapter we deal with the menaces coming from the sky while the next chapter deals with hazards due to abuses of the Earth itself.

What makes the Earth so vulnerable to the hazards coming from the sky is the fragility of its three protective shields. By decreasing distance to the Earth, the first shield is the heliosphere (Figure 3.1 [1]): a cavity in our galaxy produced by the solar wind which exerts its pressure against interstellar gas. Inside the heliosphere, the solar wind is traveling at supersonic speeds of several hundred kilometers per second in the vicinity of the Earth. Well beyond the orbit of Pluto, this supersonic wind slows down to meet the interstellar gas. At the termination shock – a standing shock wave – the solar wind becomes subsonic with a velocity of about 100 km/s. The second shield is the Earth's magnetosphere, which was already described in the previous chapter (Figure 3.2), and the third is the Earth's atmosphere itself: the thin fragile layer that allows us to breathe, that filters the lethal ultraviolet photons from the Sun and secures our survival against the most dangerous hazards of cosmic origin (Figure 3.3). Both the heliosphere and the magnetosphere act as magnetic 'shields'. They prevent the penetration of cosmic rays or divert their trajectories and, in the case of the magnetosphere, also of the

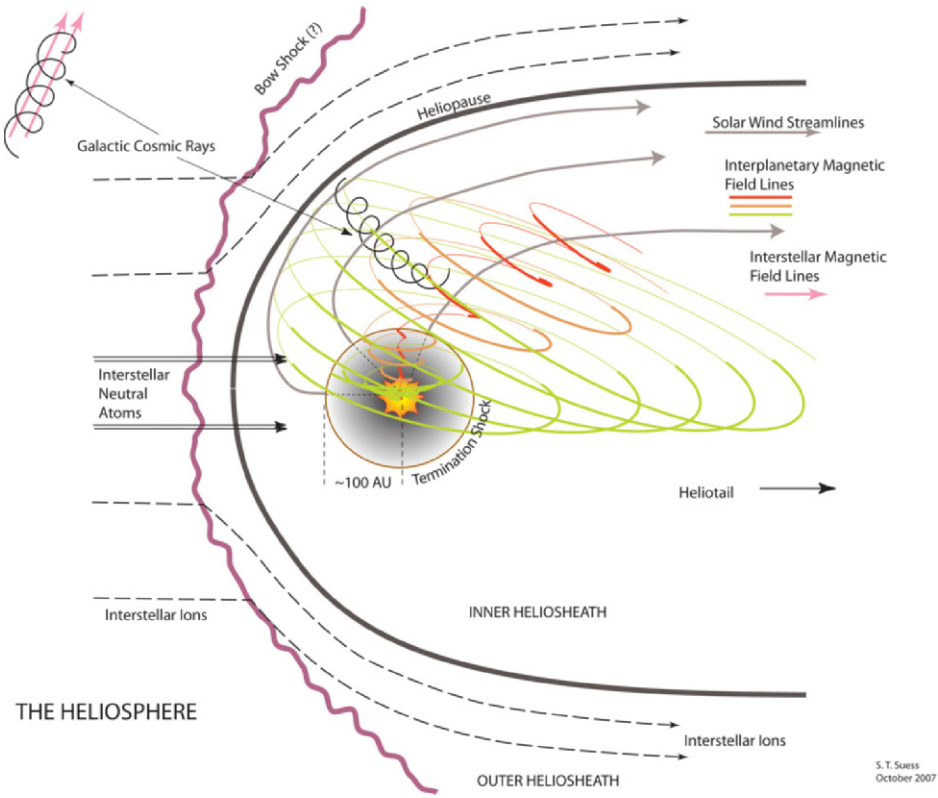


Figure 3.1 The heliosphere is a cavity of the local interstellar medium which is shaped by the magnetic field of the Sun. It is able to divert the penetration of galactic cosmic rays into the Solar System. (*Credit*: S.T. Suess, see reference [1].)

solar wind and of solar eruptions which can affect the genetic material of living organisms. The gases in the Earth's atmosphere, in particular the ozone layer, offer an efficient shield against lethal ultraviolet radiation and can at the same time 'filter' the smallest asteroids through friction. All three shields are fragile and can be affected by the different types of menace that are described in this chapter.

3.2 Galactic hazards

Observations and theory tell us a lot about our planet's 'natural' long-term future fate. The Universe contains a good hundred billion galaxies and our own galaxy contains a good hundred billion stars. It is impregnated with violence. In it, tremendous energies are at play which can trigger cataclysms, local apocalypses: collisions, shocks, explosions and bursts of light, stars collapsing in less than

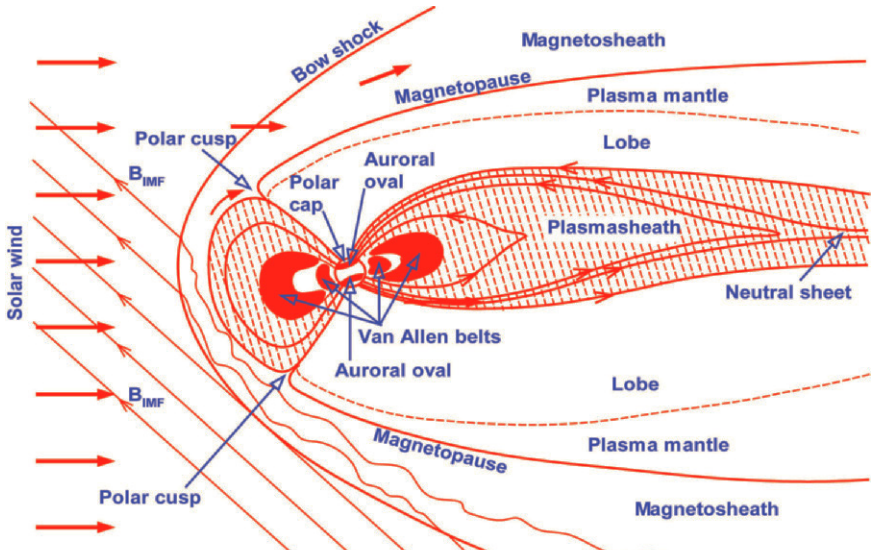


Figure 3.2 The Earth's magnetosphere constitutes a natural shield against solar particles and other cosmic radiation. Its existence is directly linked to that of the intrinsic Earth's magnetic field. Its asymmetric shape results from the magnetic pressure exerted by the solar wind which carries the solar magnetic field to the orbit of the Earth. During solar maximum, the most energetic 'gusts' of the solar wind can compress the magnetosphere down to 20,000 km from the Earth's surface. (*Credit*: A. Balogh.)



Figure 3.3 The Earth as viewed by astronauts of the International Space Station in July 2006. The Earth's atmosphere makes the Moon appear as a blue crescent floating far beyond the horizon. Closer to the horizon, the diffusion of light by the molecules of the atmosphere gradually makes the lunar disk fade away. As one looks higher in the photograph, the increasingly thin atmosphere appears to fade to black. (*Credit*: NASA–GSFC.)

1 second, black holes accreting all matter in their neighborhood, and permanent bombardment by cosmic-ray particles, dust, and pieces of rock. All of these characterize the most hostile environment we can imagine for a planet like ours.

Although a very slow process, galaxies may collide with other galaxies, and this is not exceptional since it is estimated that about 2% of the known galaxies in the Universe are observed to be in collision (Figure 3.4). In our neighborhood, Andromeda, our twin galaxy, is on a collision course with our Milky Way at a velocity of 500,000 km/h. The 'collision', which would be marked by the acceleration of the impactor as it feels more strongly the gravitational attraction of our Milky Way while approaching it, is estimated to occur in 3 billion years, largely outside our 100,000-year time frame, and does not represent an immediate menace. Furthermore, the process would be very slow and would not necessarily directly affect our Sun. On the contrary, the large clouds of gas and dust that occupy tremendous volumes inside the two colliding galaxies would feel the shocks, resulting in a large number of new stars being formed. Several of these would not last very long and would explode as supernovae. Indeed, supernovae do represent serious menaces to the neighboring stars and their planets. About 1.6 billion years after their collision, Andromeda and our Milky Way would have merged into a new single elliptical object. At that time, our Sun would be close to the end of its 'natural' life.



Figure 3.4 Spectacular collision of two galaxies: the large NGC 2207 (left) and the small IC 2163, as observed with the Hubble Space Telescope. Strong tidal forces from NGC 2207 have distorted the shape of IC 2163, flinging out stars and gas into long streamers stretching out 100,000 light-years towards the right-hand edge of the image. (*Credit*: NASA–ESA–The Hubble Heritage Team.)

3.2.1 The death of the Sun

Our Sun's life is programmed by its mass and the rate at which it burns, through fusion, its 1.4×10^{27} tons of hydrogen into helium. This is how it has been shining for 4.56 billion years and how it should continue shining for the next 6 billion years. Its brightness will increase continuously at a rate of 10% per billion years, until it exhausts its reserve of hydrogen, becoming a red giant of several hundred times its initial diameter, living then on helium fusion only. This will last for just a few hundred million years until the Sun sheds into the galaxy about one-quarter of its original mass. The rest will condense into a white dwarf as pale as the full Moon light. Long before, approximately 1 billion years from now, the Earth would have been transformed into a Venus type planet, with a Sun 10% brighter, boiling off our oceans, creating an irreversible greenhouse effect that would in the course of just a few million years transform our planet into a hot and dead body (see Chapter 9). This foretold death of the Earth, and probably of all forms of life on it, is a well-documented story. Again, the time when it will occur is far beyond our 100,000-year time frame and we should not have to fear it too much. We just keep it in mind. We should, however, worry more about some immediate hazards of pure cosmic origin.

3.2.2 Encounters with interstellar clouds and stars

Some of these hazards may result from the motion of the Solar System as it rotates around the center of the galaxy and periodically crosses its spiral arms every 150 million years [2]. These crossings, which may last a few million years, only represent a potential danger for the Solar System and its planets as the gravitational perturbations they induce on the Oort Cloud may divert comets and asteroids and send them in collision courses with the Earth (Section 3.3).

The Sun may encounter clouds of interstellar matter which are more frequent in the spiral arms of the galaxy. In fact, such an encounter may occur in about 2,000 years. It is most probable that the relatively low densities of these local 'bubbles' of gas and dust will not present a real danger, but in 2 million years the Sun may cross a more dense cloud called Ophiuchus which may be more potentially harmful, with consequences on the Earth's climate. During the 200,000 years it would take the Solar System to travel through this cloud, the Earth's atmosphere may be filled with dust, which would choke out sunlight and initiate a new glacial period [3]. Some abnormal coolings of the climate around 600 and 500 million years ago may be explained by this phenomenon.

Furthermore, the relatively fast-moving ionized hydrogen atoms and molecules in the cloud may react with the Earth's atmosphere and damage the ozone layer (see below) [4]. Normally, the solar wind would protect the Solar System against the penetration of these fast-moving particles called Anomalous Cosmic Rays (ACR) which get ionized and accelerated when they enter the heliosphere. The pressure of the ionized gases from the encountering cloud may overcome that of the solar wind, exposing the Earth to their harmful effects (Figure 3.5), one being the loss of stratospheric ozone. The high-energy particles of the ACR have enough energy to break atmospheric nitrogen and form

nitrogen oxides, generically called NO_x which destroy ozone (O_3) through the catalytic cycle of reactions:

$$NO + O_3 \rightarrow NO_2 + O_2$$
$$NO_2 + O \rightarrow NO + O_2$$

The penetration of cosmic rays may be amplified at the time of reversals of the Earth's magnetic field during which the strongly distorted magnetosphere would no longer be able to fully play its protective role. The combined effect of a cloud-crossing and of a magnetic field reversal would enhance the abundance of stratospheric nitrogen oxides 100 times at altitudes of 20–40 km, resulting in at least a 40% loss of ozone at mid-latitudes and 80% in the polar regions, exposing the Earth's surface to an increase of lethal UVB radiation [4]. A 50% decrease in ozone column density leads to an increase in UVB flux transmission of approximately three times the normal flux. The ozone loss would last for the duration of the reversal and could ultimately trigger global life extinction. The probability of cloud-crossing and magnetic field reversal to occur contemporaneously is rather low. As discussed in the previous chapter, a reversal might happen in the next 100,000 years while the crossing of the potentially dangerous Ophiuchus cloud may happen in 2 million years.

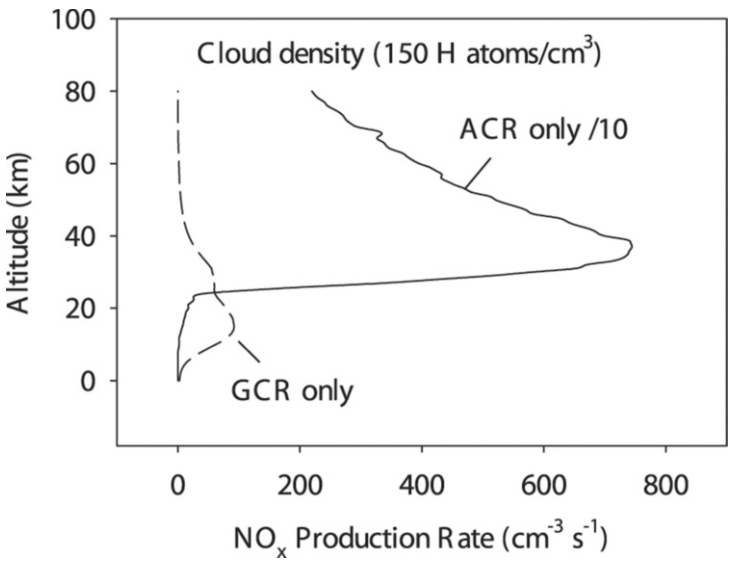


Figure 3.5 NO_x production rate by normal Galactic Cosmic Rays (GCR) and Anomalous Cosmic Rays (ACR) produced by a cloud of 150 H atoms /cm^3. The ACR production rate was divided by 10 so that it could be compared to the GCR production rate. The higher altitude of the maximum NOx production rate is due to the 'softer' energy spectrum of the ACR compared to the GCR. (*Credit*: A. Pavlov et al. [4].)

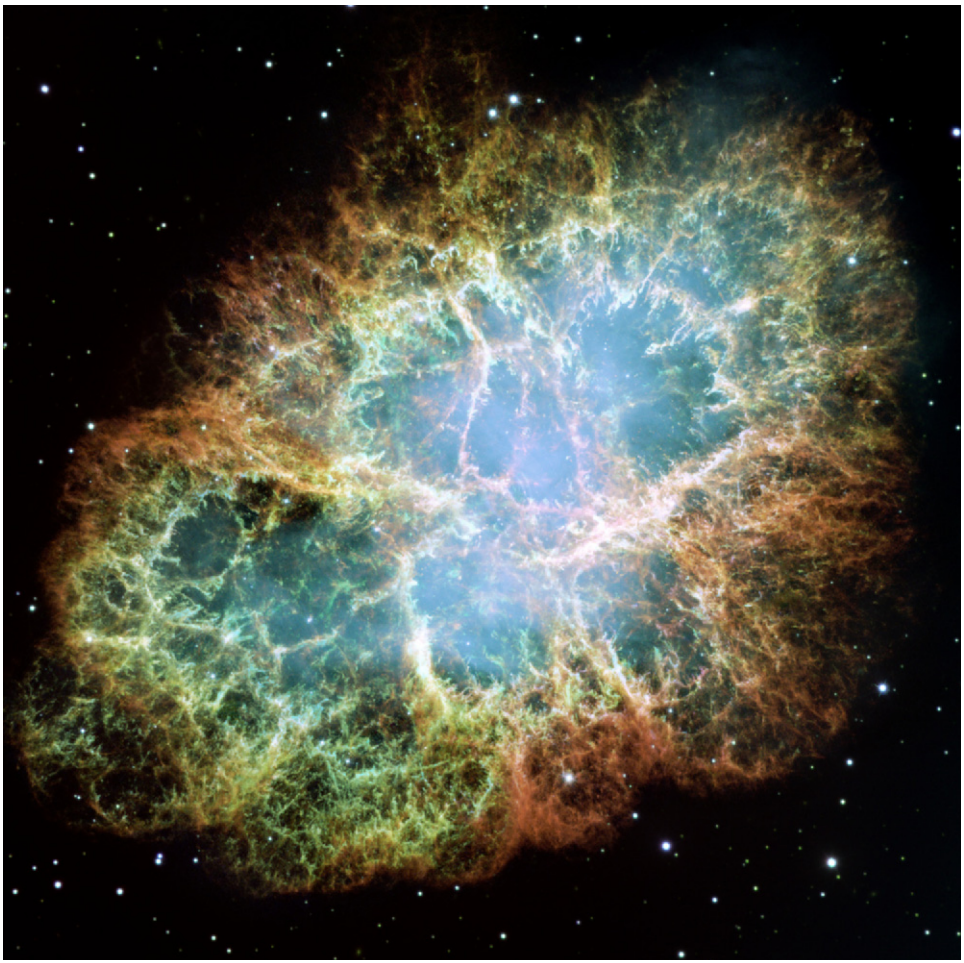


Figure 3.6 The Crab nebula was first observed in the western world in 1731 and corresponds to a bright supernova that was recorded by Chinese and Arab astronomers in 1054. Thanks to these observations, the nebula became the first astronomical object recognized as being connected to a supernova explosion. Located at a distance of about 6,300 light-years from Earth, the nebula has a diameter of 11 light-years and is expanding at a rate of about 1,500 km/s. (*Credit*: NASA–ESA and Space Telescope Science Institute.)

3.2.3 Supernovae explosions, UV radiation and cosmic rays

A second hazard is the possibility for several of the nearby stars to explode. The most dangerous are those with a mass more than 8 times the mass of the Sun. After they have burnt their nuclear fuel they collapse under their own weight. As atoms in the star's nucleus are squeezed together, they rebound outwards and end up as spectacular supernovae explosions lasting about just 1 second. The Crab nebula (Figure 3.6), which was observed by the Chinese in 1054 AD, is the

result of the explosion of a star of 10 solar masses located some 6,000 light-years away. In the weeks following the explosion, the star's brightness reached 10 billion times the Sun's brightness. These explosions are in fact the only source of production of elements heavier than iron, which are necessary for life on Earth. Such events are not very frequent: 2 to 5 per century in our galaxy which contains about 10^{11} stars. However, supernovae can often cluster in space and time.

This is the case for Scorpius–Centaurus, an association of hot young stars, located some 300 light-years away from us, which has produced about 20 supernovae explosions in the last 10 million years [5, 6]. Occurring close to the Earth, they might present a potential hazard to our planet. Their main effects are the emission of a flux of very high energy cosmic rays, and also a large increase of ultraviolet light. Both would result in the loss of the ozone layer – mostly at high latitudes – which would occur in a few minutes only but would last for several years, causing a death rate from cancer that would exceed the birth rate, and leave little chance to all living organisms to survive. Fortunately, it has been estimated that the effect of the ultraviolet flux is minimal for supernovae further away than 25–30 light-years [7]. Since we know that the Sun will not meet any massive stars so closely during the coming 100,000 years, this hazard should not concern us here. However, the menace due to cosmic rays is more real.

It has been suggested that 2 million years ago a star located at 120 light-years from Earth from a subgroup of the Scorpius–Centaurus association, the Lower Centaurus Crux, could have exploded [6]. Curiously, that time corresponds to a mass extinction of molluscs' species in tropical and temperate seas at the boundary between the Pliocene and Pleistocene. At such a relatively long distance, the main effect of the explosion would be the sudden increase in cosmic-ray flux and the subsequent destruction of the ozone layer through the increased production of nitrogen oxides [8]. The molluscs feed on sea-surface plankton which would have been damaged by the increased solar UV radiation passing freely through what looks like a precursor of the ozone hole. Cosmic rays produced in supernovae some 60 light-years away with energy densities one to two orders of magnitude higher than the average 'quiescent' value, would yield a reduction in ozone up to 20%. In support of this theory, the very unstable iron isotope, ^{60}Fe, has recently been discovered in deep Earth cores which have been deposited precisely 2 million years ago. The amount of ^{60}Fe thus found corresponds to what might be expected from a supernova exploding at a distance of about 100 light-years. The damage caused to the Earth could have lasted up to 1,000 years, a time long enough to eliminate a substantial number, if not all, of the living species.

3.2.4 Gamma-ray bursts and magnetars

Gamma-ray bursts were first discovered by US military satellites in the late 1960s and remained mysteries until 1997 when a combination of observations made in the X-ray and visible part of the spectrum allowed their origin to be understood and their distances to be evaluated. They go off at the rate of one per day all over

the sky, as compared with one supernova every few decades or so in our galaxy. Some probably correspond to a form of supernova produced by a very massive star more than 15 times the mass of the Sun, whose collapsed core is most likely a black hole. Their origin is still a matter of debate [9]. Jets formed near the black hole plough outward and accelerate to velocities very near the speed of light. The jets contain relativistic winds that interact and collide, creating shock waves and emitting high-energy cosmic rays and gamma rays. Lasting anywhere from a few milliseconds to several minutes, gamma-ray bursts shine hundreds of times brighter than a typical supernova and about 10^6 trillion times as bright as the Sun, making them briefly the brightest source of gamma-ray photons in the observable Universe.

They are all far away from us (only four have been spotted within 2 billion light-years of the Earth) because more very massive stars were formed in the early Universe than in the more recent past. But what if one were to shine up in our neighborhood? It is in fact probable that at least once in the last billion years the Earth has been irradiated by a gamma-ray burst from within 6,000 light-years in our galaxy [10]. The effects of one such impulsive burst as it penetrates the stratosphere would cause a globally averaged ozone depletion of 35%, reaching 55% at high latitudes. Significant depletion would persist for over 5 years after the burst. Additional effects include a production of nitrogen dioxide, NO_2, whose opacity in the visible would lead to a cooling of the climate over a similar timescale. These results support the hypothesis that a gamma-ray burst may well have initiated the late Ordovician mass extinction 443 million years ago, which coincided also with times of high CO_2 concentrations.

In the mid-1990s a new source of intense gamma-ray radiation was discovered called 'magnetar'. Magnetars are thought to be the remnants of supernovae but, in addition, they possess a magnetic field with the strongest intensity ever observed in the cosmos, equal to some 10^{15} times the Earth's surface field, hence their name. They throw out bursts of gamma-ray and X-ray radiation lasting a fraction of a second with an energy equivalent to what the Sun emits in an entire year! The temperature of the plasma emitting the burst has been estimated at 2×10^9 K. One such object, called SGR 1806–20, was observed in 2004 at a distance of almost 50,000 light-years in the Sagittarius constellation. Its burst was the brightest ever observed and caused problems for several satellites. Also, the Earth's ionosphere and radio-communications were affected. It was probably caused by a sudden readjustment of the huge magnetic field anchored in the neutron star which underwent a monstrous 'star quake', releasing a substantial quantity of the internal energy stored in the field.

Magnetar bursts are less energetic than gamma-ray bursts, but they occur more frequently and are more likely to happen close to the Solar System. So far, just a dozen magnetars have been found, two of them in our galaxy. If one were to appear at 10 light-years or closer to us, its high-energy radiation (gamma rays, X-rays and ultraviolet) would significantly deplete the ozone layer. It is impossible at this stage to give any number as to their frequency of occurrence which increases for magnetars located at greater distances. We may guess that

Table 3.1 Average number of cosmic hazards of galactic origin per unit time

Supernovae	< 25 light-years	1 per 3 billion years
Gamma-ray burst	< 6000 light-years	< 1 per billion years
Magnetar	< 10 light-years	1 per 5 billion years (?)
Magnetar	< 60 light-years	10–100 per billion years

one might have occurred already in the lifetime of the Sun at a distance smaller than 10 light-years.

Table 3.1 gives an estimate of the frequency of the known galactic hazards just described. They may have had some catastrophic effects during geological times, and similar ones may not be excluded during the coming 100,000 years. Gamma-ray bursts cannot be predicted but the chance of one occurring during that period is no more than 1 in 10,000 or even less. However, as frightening as they appear, these threats are nothing as compared to the violence that hangs upon us at a much shorter distance.

3.3 Solar System hazards

From the beginning of their existence, the planets of the Solar System have been bombarded by meteorites and rocks through the accretion process with increased intensity during the Late Heavy Bombardment. As mentioned in Chapter 2, one asteroid impact occurring 65 million years ago creating the Chicxulub crater in the region of Yucatan in Mexico was most probably the cause of the sudden disappearance of all dinosaurs on Earth. The size of the impactor was about 10 km. The chances of being hit by an asteroid were certainly higher in the early age of the Solar System than now. However, the probability for such an event to occur again now or in the future is not negligible as the bombardment continues, but with more moderate intensity. So much so that scientists, space agencies, politicians and now the United Nations are actively involved in establishing the probabilities of potential future impacts and in studying mitigation strategies in view of protecting humanity from their potentially deadly consequences. One of the most comprehensive discussion about hazards due to asteroids and comets can be found in Gehrels [11].

3.3.1 Past tracks of violence

Watching the Moon with a pair of simple binoculars reveals a telling landscape. There are craters everywhere! Devoid of an atmosphere, the Moon accumulates all past tracks of violence. It is in fact beating the record in the Solar System of the biggest crater, the Aitken basin near the lunar South Pole, with a diameter of 2,500 km and a depth of 12 km (Figure 3.7). Craters are also visible on the surfaces of all the solid bodies including the satellites of all planets that we have been able to watch up to now with space probes. The asteroids themselves are not spared from collision accidents (Figure 3.8).

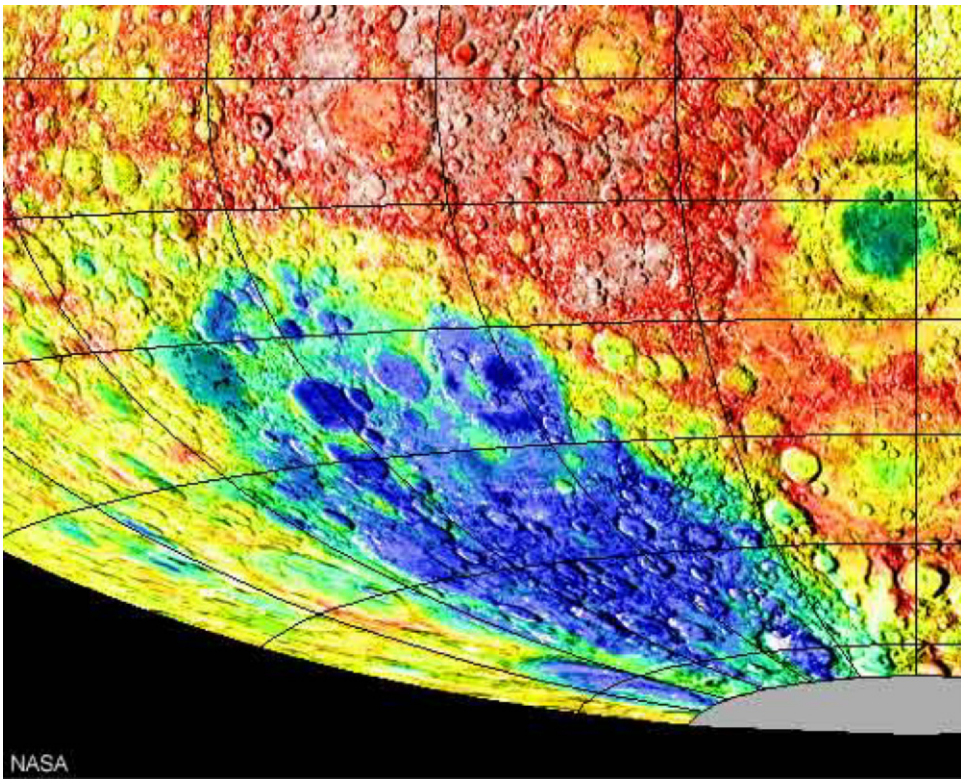


Figure 3.7 Much of the area around the Moon's South Pole is within the Aitken basin shown in blue on this lunar topography image, a giant impact crater 2,500 km in diameter and 12 km deep at its lowest point. Many smaller craters made by later impacts exist on the floor of this basin. (*Credit*: NASA/National Space Science Data Center.)

Even on Earth, signatures of the most energetic impacts can still be found. Unfortunately, our planet is not very helpful in that respect because it smoothes out the scars left by the impacts, as a consequence of plate tectonics, volcanism, wind and water erosion, sedimentation, etc. According to the Earth Impact Database in Canada, about 170 craters have been inventoried on Earth. Obviously, it is easier to identify the most recent impacts than the old ones whose craters and debris are buried under sea water and sediments. The oldest has a diameter of 250–300 km and is some 2 billion years old. It is located in South Africa, in the region of Vredefort, 110 km south-west of Johannesburg. It was most likely caused by a 10-km asteroid, similar in size to the object that formed the Chicxulub crater (Figure 3.9). The most recent one is the Meteor Barringer crater in Arizona whose diameter of 1.3 km is the result of an impact 50,000 years ago by a small nickel and iron asteroid of about 50 meters. At a velocity of 65,000 km/h, its energy was equivalent to 20 megatons of TNT or 1,300 times the strength of the Hiroshima bomb.

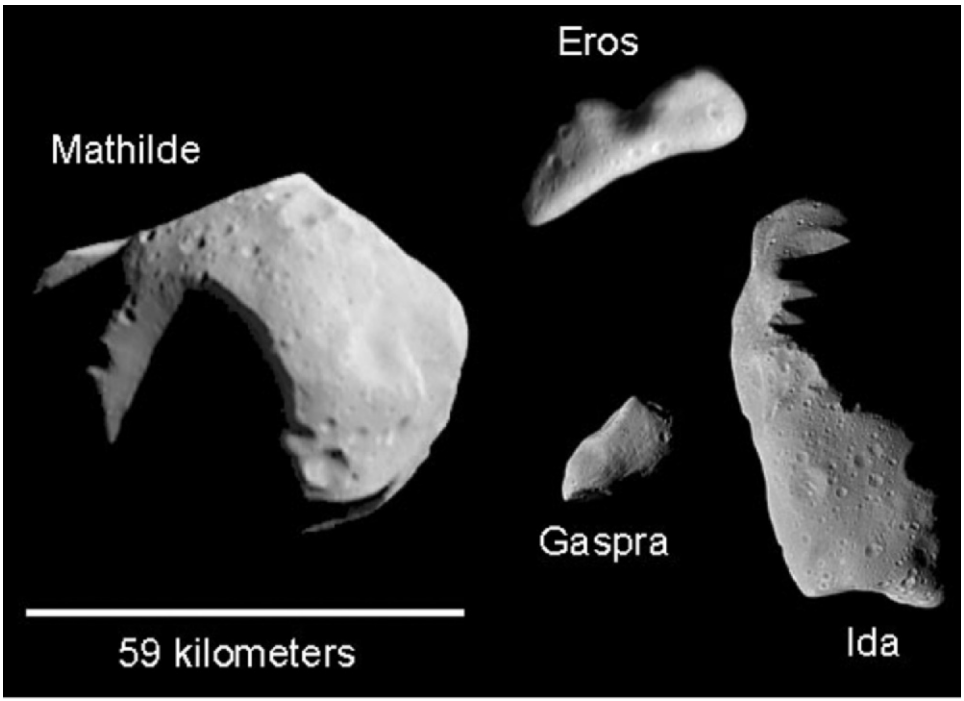


Figure 3.8 Sample of asteroids that have been explored with space probes: Eros and Mathilde by the NEAR-Shoemaker spacecraft, Gaspra and Ida by the Galileo probe to Jupiter. Mathilde, a Carboneous asteroid, is a very dark object (albedo, 3%) whose brightness has been artificially enhanced several times on this picture to match the other three. Eros is a banana-shaped body of 31.6 km × 11.3 km × 8.7 km, probably the end product of a huge collision. (*Credit*: NASA.)

Space observations do offer a powerful means for detecting impacts and understanding what conditions determined their occurrence. They are particularly interesting in areas whose geological history is more favorable for the preservation of impact tracks that are not easily accessible. For example, two twin impacts of a diameter of 6.8 and 10.3 km respectively, estimated to be 140 million years old and formed by a pair of meteorites of approximately 500 meters, have been discovered in south-east Libya using optical and radar-imaging techniques (Figure 3.10) [12].

A meteoritic impact liberates an enormous amount of energy depending upon the size, density and velocity of the impactor. These velocities range between 11.2 km/s (the escape velocity of the Earth–Moon system) and 72 km/s (the orbital velocity of the Earth plus the escape velocity of the Solar System at the distance of 1 Astronomical Unit (AU)). The collision of a 10-km object with the Earth would locally raise the temperature to several thousand degrees and the atmospheric pressure in the resulting shock wave nearly a million times. This is equivalent to several tens of millions of megatons of TNT. The effect is not just

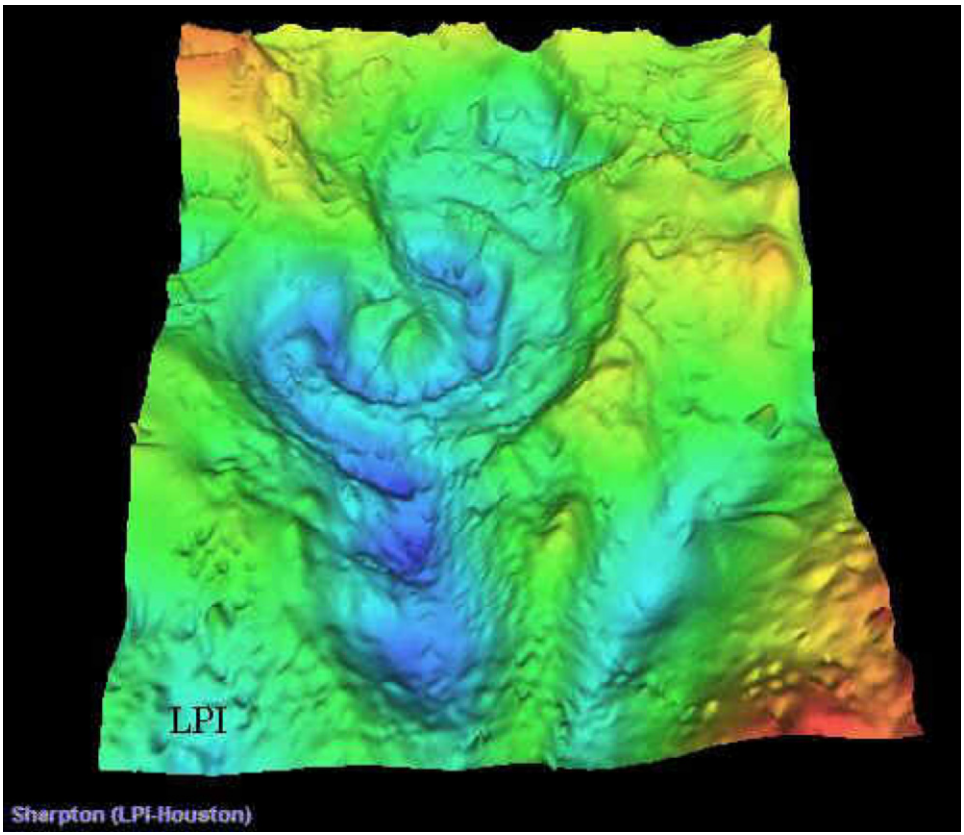


Figure 3.9 Three-dimensional map of the 180-km Chicxulub crater in Yucatan, Mexico, obtained through seismology techniques using the combined reflection/refraction of seismic waves artificially generated, revealing the details of the topography of the impact. (*Credit*: V.L. Sharpton and the Lunar Planetary Institute.)

the crater and the local destruction, but also the erosion of the ozone layer by the chemical species released in the atmosphere by the intruder and the complete alteration of the global climate for a very long time.

The Chicxulub meteorite was particularly deadly. One reason – which may leave some hope that a similar size object may not cause the same degree of devastation in the future – is that the level of damage seems to depend to a large extent on the mineral composition of the ground at the impact point. The presence of carbonates and sulfates, which cover only 2% of the Earth's surface, is particularly crucial in determining whether the collision will be devastating or not. It is exactly the case for the Chicxulub impact which vaporized rocks made of such compounds, pouring carbon and sulfur dioxide into the atmosphere. On the contrary, the Popigai crater in Siberia, one of the fourth largest in the world, formed 35.7 million years ago by a large body comparable to the Chicxulub

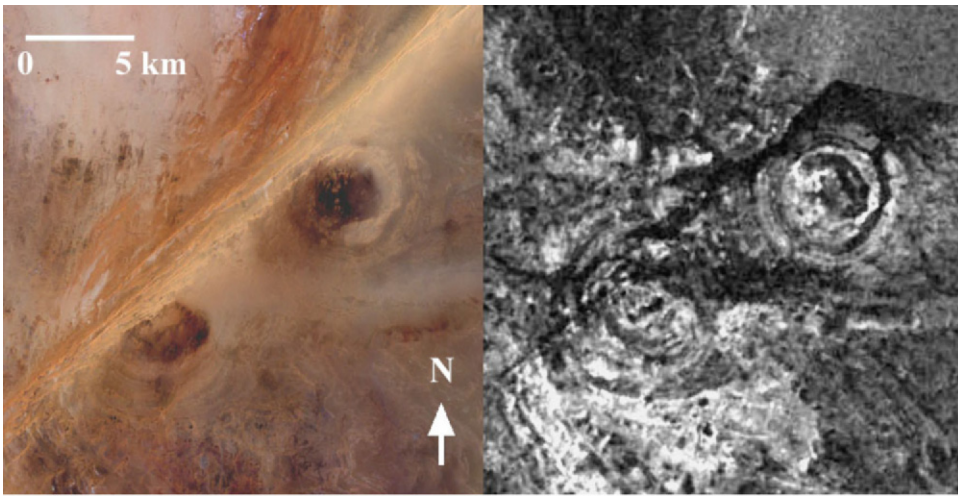


Figure 3.10 Landsat image of a double impact crater (left) and the corresponding JERS-1 L-band radar image (right) at a resolution of 100 meters. (*Credit*: P. Paillou, reference [12].)

meteorite, is not associated with any noticeable contemporaneous major extinction.

3.3.2 The nature of the impactors: asteroids and comets

The Solar System contains trillions of asteroids and comets. This number was evaluated through statistical analysis which unfortunately does not allow all of them to be inventoried. In fact, only about 300,000 of those have been reported. Hence, it is not clear which of them present a real danger. They are different in nature, originating from different places in the Solar System (Figure 3.11(a) and 3.11(b)). Their sizes vary considerably from object to object, from a few microns for dust grains to a few hundred kilometers. The small undifferentiated objects are thought to be the most primitive bodies around in the Solar System. Some may have aggregated from dust; some may be the by-products of shocks among themselves.

Beyond Pluto's orbit, between 30 and nearly 50 AU, more than 1,000 objects have been detected, 60 of which have a diameter of 100 km or larger, forming the Kuiper Belt. They are debris left over from the building of the outer planets or pieces that could not make a planet because of the perturbations induced by Jupiter. Between 5 and 30 AU, gravitational interactions by Saturn and Jupiter have emptied interplanetary space from asteroids. Between Mars and Jupiter lies the Main Asteroid Belt (MBA) which contains about 2 million objects. There are probably more than several tens of thousands with a diameter over 1 km, with some over 100 km. Ceres, the largest ever discovered, is roughly 913 km in diameter, Pallas 523 km, Vesta 501 km and Juno 244 km. Their orbits are in the ecliptic plane, but under the combined effects of collisions – mostly among

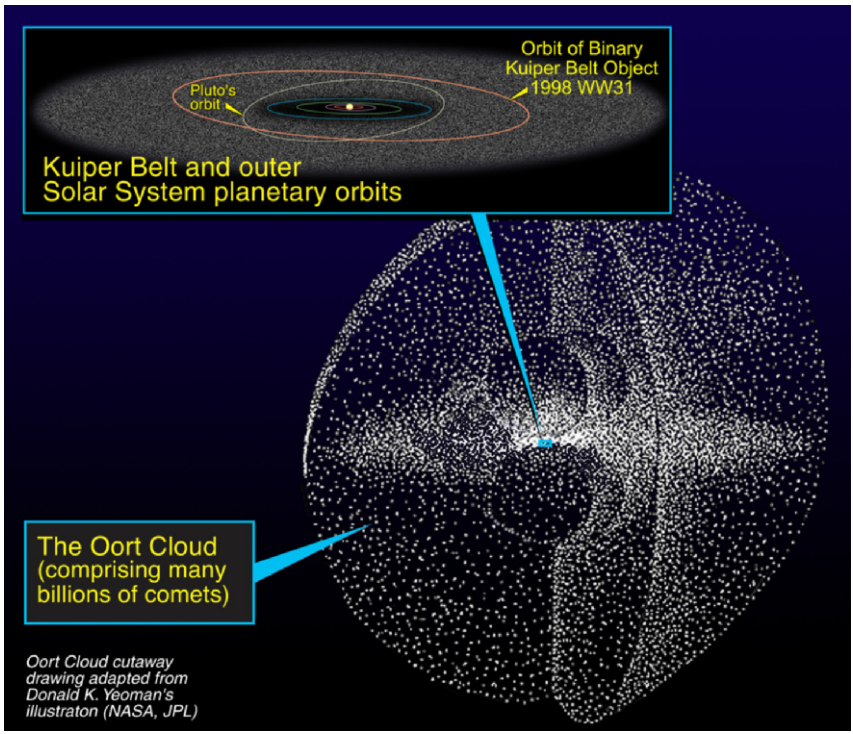


Figure 3.11(a) The Oort Cloud extends from 50,000 to 100,000 AU from the Sun and contains billions of comets. The Kuiper Belt extends from the orbit of Neptune at 30 AU to 50 AU. The objects within the Kuiper Belt, together with the members of the scattered disk extending beyond, are collectively referred to as trans-Neptunian. The interaction with Neptune is thought to be responsible for the apparent sudden drop in number of objects at 48 AU. (*Credit*: NASA–JPL.)

themselves – and of the attraction of either Mars or Jupiter, their inclination might change. Some may retire far away in the vicinity of Saturn or even beyond the orbit of Neptune. The Chicxulub meteorite may have detached from its orbit as the consequence of a huge wobble that affected the whole inner Solar System around 65 million years ago, and may also have altered the orbits of Mars, the Earth, Venus and Mercury as well as the asteroid belt [13]. The origin of the killer has now been traced back (with a probability >90%) to a 170-km-diameter object that broke up some 160 million years ago in the inner Main Belt Asteroids whose fragments slowly migrated by dynamical processes to orbits where they could eventually strike the terrestrial planets [14].

The so-called Near-Earth Asteroids (NEAs) have orbits that pass within 45 million km from that of the Earth with a distance to the Sun smaller than 1.3 AU. They represent a distinct population from normal Main Belt Asteroids that move in the region between Mars and Jupiter. Most of the NEAs appear to be true asteroids; however a small number have highly eccentric orbits and are probably

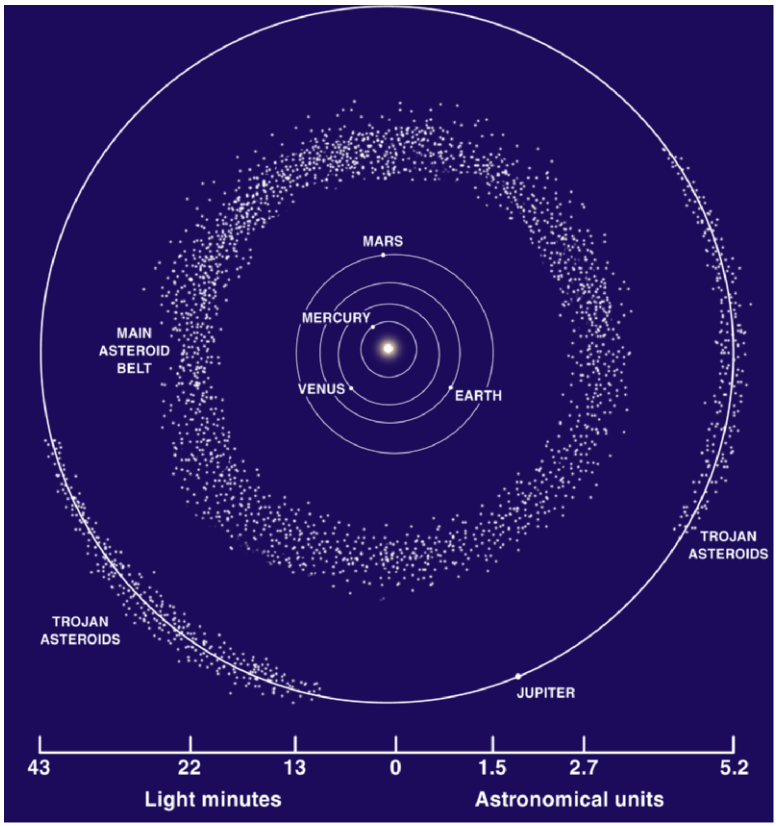


Figure 3.11(b) The Main Asteroid Belt is between the orbits of Jupiter and of Mars. The group that leads Jupiter on its orbit is called the 'Greeks' and the trailing group on that orbit is called the 'Trojans'. (*Credit*: NASA-JPL.)

extinct comets that have lost all their volatile constituents. Tens of thousands probably exist but only about 5,000 were known as of December 2007 (Figure 3.12), ranging in size up to ~32 km. Approximately 1,000 of these objects measure about 1 km, and, currently, about 850 of them have minimum distances between their orbits and the Earth that are smaller than 0.05 AU (about 7,500,000 km). These are known to be potentially hazardous objects (PHOs). It is estimated that there could be over 100,000 asteroids and comets, including 20,000 PHOs, once the smaller 140-meter and larger objects are added to the catalog. New observations are definitely required to refine these numbers. By use of a dedicated space-based infrared system and optical ground-based observatories, NASA has received a mandate by the US Congress to detect and identify by 2020 at least 90% of all the 20,000 estimated potential killers. The observations will be gathered by the Jet Propulsion Laboratory and the Near Earth Object Dynamic System of the University of Pisa in Italy.

There exist several types of asteroid depending upon their composition: C-

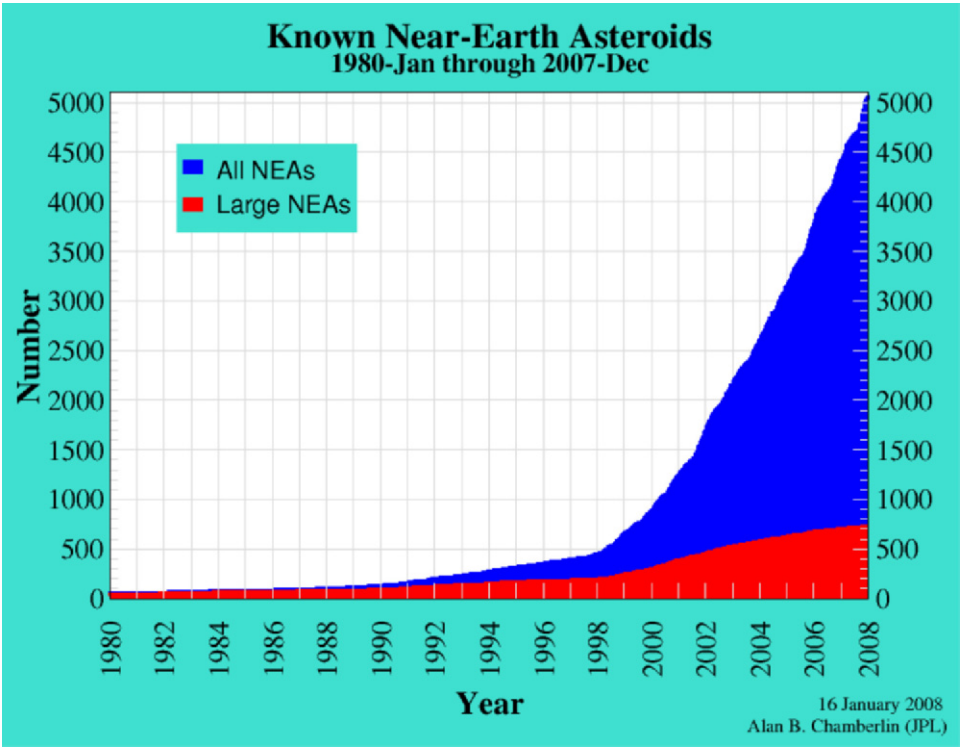


Figure 3.12 Known Near Earth Asteroids from January 1980 to December 2007. The blue area shows all the known NEAs and the red only those larger than 1 km. (*Credit*: NASA–JPL.)

type (or carbonaceous) represent 75% of the known population; S-type (silicaceous), 17%; and M-type (metallic), 8%. About 15% are associated and form binary systems. Most of the NEAs above 200 meters are dark, their surface being covered with a blend of grains, stones in the millimeter and centimeter range, and rubble, usually called regolith. A few are solid iron. We know the mass of only 20 of the NEAs, because we do not know precisely their internal structure, their density, their albedo – the proportion of the Sun's light they are able to reflect – and therefore their dimensions. Sizes and masses are indirectly inferred and may be wrong by factors of ~ 2 and ~ 10 respectively. Most objects smaller than 100 meters are monolithic, while some may be less cohesive and more fragile. Much remains to be done to properly evaluate their properties especially if they are considered to be potentially dangerous. Recent progress has, however, been made in that respect, in particular by NASA and by the Japanese.

For the first time in the history of space exploration, in February 2001, NASA landed its Near Earth Asteroid Rendezvous probe on the surface of Eros, an S-type asteroid orbiting the Sun between 1.13 and 1.78 AU, with a dimension of 31.6 km $\times$ 11.3 km $\times$ 8.7 km (Figure 3.8). The scientific data collected in the course of the mission have confirmed the compact rocky structure of Eros, made up

internally of fragments of original materials from the solar nebula composed of approximately 10% of iron and nickel. More than 100,000 craters larger than 1.5 meters were found on its surface, a density close to saturation which evidences what a hectic life Eros had (!), as it was submitted to an intense bombardment in the course of the last 1 to 2 billion years. The largest of these craters, poetically named Psyche – probably the result of an impact caused by a projectile of about 300 meters traveling at 5 km/s – has a diameter of more than 5 km and is 900 meters deep. On such a small body, the shock must have been dramatic and has most likely caused the asteroid to change orbit and fall into the gravitational attraction of Mars.

The double landing of the Hayabusa Japanese probe – Falcon in English – on the 500-meter class S-type asteroid 25143-Itokawa on 19 and 25 November 2005, was to return samples collected from the surface. (Owing to an improperly timed set of commands, it was still not certain at the time of printing this book that samples had actually been collected.) Itokawa crosses the orbits of Earth and Mars in a 1.5-year orbit around the Sun [15]. With the help of its ion engine, Hayabusa is now on its way back to Earth, where it may drop the sample capsule into the Australian desert in June 2010. Hayabusa observed Itokawa's shape (540 $\times$ 270 $\times$ 210 meters) and measured its geographical features, its reflectance, mineral composition, and gravity, from an altitude of 3 to 20 km (Figure 3.13). Rather than a solid piece of rock, these observations revealed a surprisingly fragile body resembling a pile of rubble held together by the weak gravity of the asteroid, covered with regolith presenting also some very smoothed areas where the dust has most likely migrated towards the poles as a result of vibrations induced by the impacts of meteorites [16].

As far as comets are concerned, the majority of them are located at the outskirts of the Solar System in the Oort Cloud. When they approach the Sun, they are much easier to see than asteroids, even with the naked eye. However, in the cold of deep space, before they develop their tail – also called a coma – they are as dark as their rocky brothers. Long-period comets are potentially dangerous and come as a surprise from the Oort Cloud. From time to time, when the Sun passes in the vicinity of other stars, their gravity may perturb the orbits of the comets and in rare cases send them on a possible collision course with the Earth. Their number is estimated to be several billion but they probably account for only about 1% of the impacts. Periodic comets, such as Halley**,** originating from the Kuiper Belt, are not the most dangerous. As they regularly return to the vicinity of the Sun, their orbit parameters are well known. At periodic intervals, they may be perturbed by the passage of the giant planets and get closer to the Sun, with a return time less than 200 years.

Scientists have been intrigued by comets for many years. Before the space age, it was not clear whether they possessed a rocky nucleus, or whether their structure was more like a set of sand grains stuck together. It had been suggested that they were 'dirty snow-balls' of dust and water-ice. During the night of 13–14 March 1986, the mystery was solved by Giotto, the first interplanetary mission of ESA. Giotto encountered Halley's Comet when it was 1.5 AU away, at a nucleus

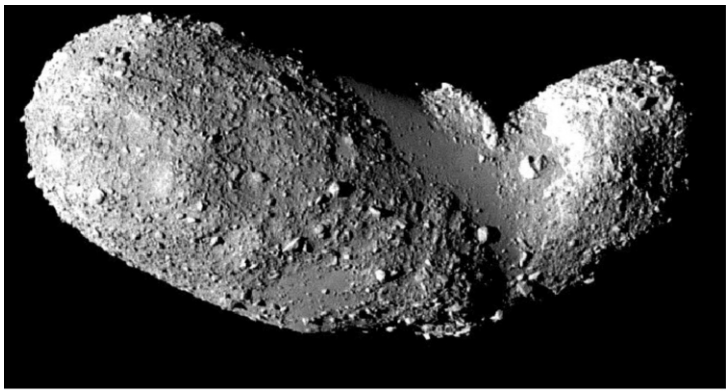


Figure 3.13 Asteroid Itokawa as observed by the Japanese Hayabusa space mission in September 2005 from a distance of 20 km. (*Credit*: JAXA-ISAS.)

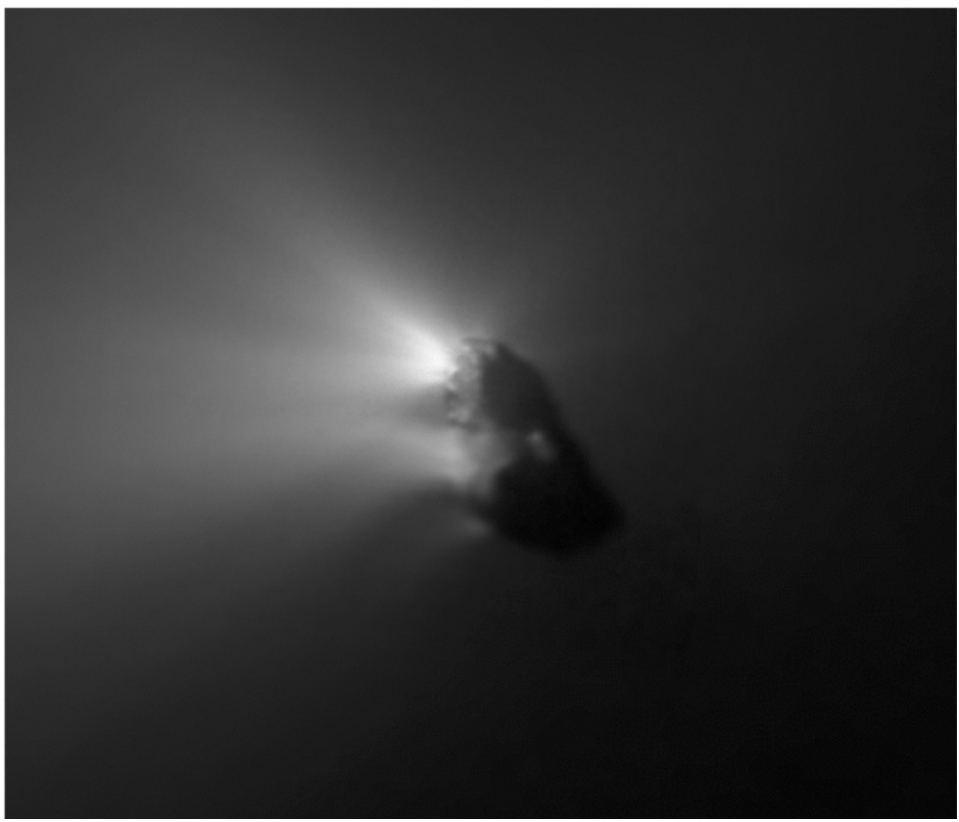


Figure 3.14 The nucleus of Halley's Comet as observed by the Giotto spacecraft on 14 March 1986 from a distance of 1,500 km. The nucleus longer dimension is about 15 km, equivalent to the size of the island of Capri west of Naples in Italy. (*Credit*: MPAE, Lindau and ESA.)



Figure 3.15 Premonitory revenge against potential killers? The objective of NASA's Deep Impact mission was to 'bombard' Comet Tempel 1 14-km nucleus by a 370-kg projectile made in USA on 4 July 2005, creating a 30-meter crater, in an attempt to analyze its internal structure. The impact released some 10,000 tons of dust and water-ice, confirming the fluffy nature of the nucleus. (*Credit*: NASA.)

miss distance less than 600 km and a relative velocity of some 240,000 km/h. The first-ever pictures of a comet nucleus at such a short distance were obtained on that night (Figure 3.14), [17]. They revealed what was then identified as the darkest object of the Solar System, containing mostly dust and water-ice representing 80% by volume of all of the material thrown out by the comet [18]. NASA's Stardust mission brought back to Earth particles collected from the nucleus of Comet Wild-2 back in 2006. They contain materials that appear to have formed over a very broad range of solar distances and perhaps over an extended time range.

Giotto and NASA's Deep Impact mission (Figure 3.15) have revealed that comet nuclei are not like hard rocks but more like fluffy objects made of ice and powder-size particles weakly agglomerated into something with the consistency of a snow bank. By their very nature, they are obviously more fragile than asteroids. When one of them encounters the Earth and plunges into its atmosphere, it may not withstand the tremendous tensions generated by its supersonic velocity, and shock waves may break it into pieces before it reaches the Earth's surface, in the same way as Comet Shoemaker–Levy broke into a myriad of 21 fragments when it came close to Jupiter in July 1994 and was torn apart by the gravity of the giant planet (Figure 3.16).

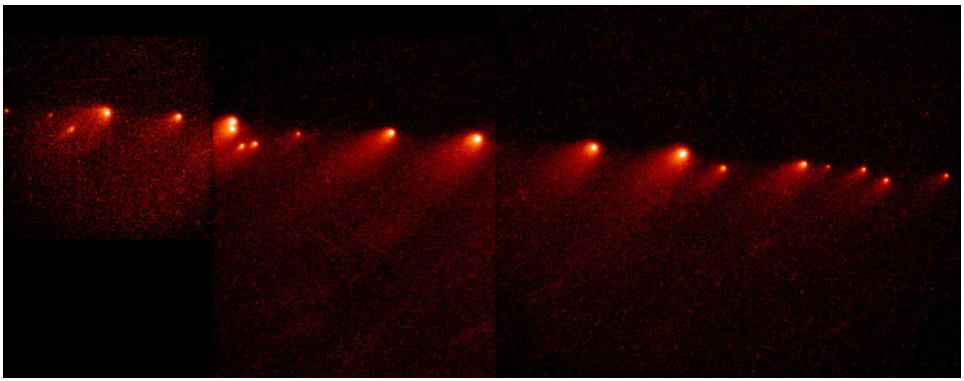


Figure 3.16 From 16 July to 22 July, 1994, due to the gravity field of Jupiter, Comet Shoemaker-Levy 9 fragmented into some 21 pieces with diameters estimated at up to 2 km. The fragments eventually collided with the giant planet at 60 km/s. This was the first collision of two Solar System bodies ever observed. (*Credit*: NASA and the Hubble Space Science Institute.)

This is most likely what happened in the Tunguska in Siberia in 1908: there was no trace of any impact, just the signature of a tremendous explosion equivalent to 15 megatons of TNT, which occurred between 6 and 8 km above ground, generating a blast wave and a high-speed wind that leveled all the trees over an area more than 2,000 km^2 and killed the reindeers that had the unfortunate idea of being present there. The size of the impactor has been estimated to be just about 50 meters across. The noise of the explosion could be heard within a radius of more than 800 km. The air over Russia, and as far as Western Europe, was filled with a fine powder of dust which stayed there for over two days. The low density of the population probably explains the reason why no casualties were reported. If the accident had occurred in a densely populated area, the situation would have been worse and the death toll catastrophic. Had it exploded above Brussels, the whole city and its neighborhoods would have been totally destroyed, resulting in a million victims.

3.3.3 Estimating the danger

The danger presented by comets and asteroids depends obviously on their size. Atmospheric friction and shock waves burn or break into pieces the fluffiest smaller than 50 meters. Such objects might create Tunguska-type problems and are locally harmful, but not globally. More dangerous are the bigger, more massive and more robust objects. With the number of these already identified, we can evaluate the proportion that are likely to hit the Earth at a given time.

Table 3.2 Fatalities estimated for a wide variety of different impact scenarios

A global catastrophe is defined as one resulting in the loss of 25% of the world's population. Above a given threshold, a local event may become a global impact catastrophe as the local effects are augmented by global climatic perturbations. A threshold is not very sharp and there exist a large range of uncertainty: for example, tsunamis produce more than local but less than global effects. (Adapted from Chapman and Morrison [19])

Type of event	Diameter of impactor	Energy (million tons)	Average fatalities per impact	Typical interval (years)
High atmospheric break-up	<50 m	<9	Close to zero	Frequent
Tunguska-like events	50 m to 300 m	9–2,000	5,000	250
Large sub-global events	300 m to 1.5 km	2,000–2.5×10^5	500,000	25,000
Large sub-global events	300 m to 5 km	2,000–10^7	1.2 million	25,000
Low global threshold	>600 m	1.5×10^4	1.5 billion	70,000
Nominal global threshold	>1.5km	2×10^5	1.5 billion	500,000
High global threshold	>5 km	10^7	1.5 billion	6 million
Dinosaurs killer type	>10 km	10^8	5 billion	100 million

Table 3.2 [19] summarizes the estimated fatalities for a large variety of sizes. The smaller objects could have severe consequences on localized areas. On the contrary, impacts from objects more than a few kilometers, although very rare, would cause massive extinction over the whole planet. Even though the probability of an event equivalent to the Cretaceous extinction has an estimated repetition time of 100 million years, it should not be concluded too hastily that we can live without fear for the next 100,000 years. Catastrophes like this might well occur at any time: tomorrow, or in the next century, or in 200 million years. To place things in correct proportion, however, the risk for a given human being of being killed by an asteroid is the same as dying in a plane crash, assuming one flight per year.

Contrary to other cosmic or natural hazards, NEO impacts can be forecasted rather easily provided we can detect the potential impactor early enough and know its orbit accurately. Systematic observations and surveys offer the only possible early warning capability to detect the most dangerous of them. Approximately 10,000 NEOs are supposed to have a non-zero probability of impacting the Earth over the next 100 years [20]. The danger depends on where on Earth the impact will occur.

Because the Earth's surface is 70% covered with oceans, they will most likely fall into water, generating gigantic tsunamis. It has been estimated that a 200-meter object could set off waves 10 to 20 meters high in the deep ocean which could grow an order of magnitude when approaching the coastline and drown all

the inhabitants of the littorals thousands of kilometers from the impact point (Chapter 4). In addition to the large number of casualties, which may reach several tens or hundreds of millions, the economic consequences of such a disaster could be in the range of several hundred billion US dollars [21]. The damage could be minimised if given sufficient warning, as the populations could be moved to more elevated areas. The degree of success of this measure is to be modulated according to the size of the object, its composition, its velocity and its angle of approach, and of course also the size of the population living in the coastal zone. The damages from small objects less than 50 metres or so can be reduced through proper civilian protection and early warning, but not completely eliminated. For the bigger objects the damage will probably last for several years, if not centuries. The effect might be intense cold as the Sun's light will be blocked by dust and debris, and subsequent infernal global warming due to the greenhouse gases released in the atmosphere by the impact once the sky becomes clearer, as most likely happened to the dinosaurs. The success of any survival operation will depend upon the possibility of sustaining a large part of the Earth's population in conditions where the light from the Sun is blocked, where the temperature is freezing cold and then becomes unbearably high.

Detecting, observing and tracking the potential impactor early enough at the largest possible distance within 1.3 AU or further can be done from the ground using optical telescopes with wide fields of view as well as radars. The smaller the objects, the larger the telescope required to detect them. Conversely, the biggest objects need only modest size telescopes: a 1-meter optical telescope is sufficient for detecting objects of a few hundred meters or more, and is hence within affordable costs. The survey must be continuous and systematic. This is why the detection systems must be fully automatic especially if they are located in remote places. Radar-type devices include the 300-meter Arecibo radio-telescope in Puerto Rico, together with the 70-meter Goldstone tracking station of NASA's Deep Space Network operated by the Jet Propulsion Laboratory in California. They form a very powerful all-weather, day-and-night system for early reconnaissance of $\sim$>300-meter asteroids and for refining the characteristics of the NEO's orbit for objects close to the Earth (< 0.25 AU). Nearly half of these systems are operated by the USA, the other half being mostly in Europe, Russia and Japan [22].

Whole sky surveillance may be easier from space, but is certainly more expensive. Artificial satellites offer very powerful capabilities because they can be constantly on the alert, operating day and night, 365 days each year. Because asteroids and comet nuclei are dark, they absorb solar light easily: their temperature increases through this process and they re-emit a substantial amount of the absorbed radiation in the infrared, which makes it possible with the appropriate instrumentation to detect them more easily in this range of wavelengths. Since the Earth's atmosphere absorbs infrared radiation, such measurements can only be conducted from space and a dedicated infrared space telescope would be an optimal instrument for early warning and characterization. The cost of such systems is, however, high and this explains why very few,

such as the ESA Infrared Space Observatory or the NASA Spitzer satellite, have been built in the past.

The portion of the sky that lies between the Earth's orbit and the Sun is difficult to observe because the light from the Sun blinds the telescopes. Objects in that region might be observed better at sunset or sunrise, provided that the proper instruments are in operation. The ideal would be to have an orbiting telescope between the Sun and the Earth (see Box 3.1). Although it is not its primary objective, ESA's Beppi Colombo mission to study Mercury, the closest known planet to the Sun at only 40% the Sun–Earth distance, if it is not canceled, will be in a unique position to observe such objects located between the orbit of Mercury and Earth. Also, not specifically dedicated to that objective, ESA's GAIA astrometry mission, designed to detect the position and the motions of stars and of all other objects in the sky with an accuracy of more than 1,000 times better than what can be done from the ground, will also observe objects very close to the Sun and be able to spot the potentially harmful NEOs with a precision 30 times better than any other telescope. The success of these missions may offer a test for the setting up of a future space observation system.

Box 3.1 SOHO

In 1995, ESA and NASA launched SOHO, an observatory looking at the Sun without interruption from a 1.5-million-km distance to the Earth, the so-called Lagrange point L_1 where the attraction of the Sun counterbalances exactly that of the Earth–Moon system. From such a unique vantage point, SOHO has revealed an unforeseen and unique capability of detecting comets when they get close to the Sun. On average six new comets per month are thus discovered, totaling more than 1,000 over 10 years. Amateurs, carefully analyzing the pictures and movies of the solar corona that are regularly shown on the web, have detected most of them. Very few could be observed previously. None of them, however, is yet considered to be a threat to us because of their small sizes. They are probably the fragments of a bigger comet that disintegrated in the vicinity of the Sun.

Once a new NEO is discovered, its trajectory is estimated, which gives a first, coarse, indication of the potential risk of the object approaching and crossing the Earth's orbit. The second step is to refine the accuracy of the prediction and the detailed observation of the object's dynamics and its physical properties. It may well be that the risk increases in scale; however, in general, the tendency is for the risk to diminish while more accurate observations are acquired.

3.3.4 The bombardment continues

Meanwhile, the bombardment keeps going. Every day, several hundred tons of cosmic dust fall on Earth, and some 50,000 meteorites every year, that are

fortunately too small to represent a real danger. More and more impactors are observed as a result of the increased observing capacity. Below is just a subset of some of the most recent observations. It illustrates both the progress in observation techniques and explains the growing concern as more and more objects are observed.

A 1-km asteroid called 1950 DA (see Box 3.2) was observed on 23 February 1950 at a distance of 8 million km. It could be followed in the sky for over 17 days until it disappeared from sight. It was observed a second time just on the eve of the 21st century, on 31 December 2000, and for that reason got the name 2000 YK 66! These two consecutive apparitions have made it possible to evaluate its size as 1.1 km and to precisely compute its trajectory, predicting its next closest visit as March 2880, with one chance in 300 that it will hit our planet and devastate as much as one full continent. It has the highest value of 0.17 on the Palermo scale (see Box 3.3 [23]).

Box 3.2 Asteroids designation

After discovery, when their orbits are not yet precisely known, asteroids generally receive a provisional designation. After its orbit is precisely known, the asteroid is given a number and finally (optionally) a name. The first element in an asteroid's provisional designation is the year of discovery, followed by two letters and, optionally, a number. The first letter indicates the half-month of the object's discovery within that year – 'A' denotes discovery in the first half of January, 'D' is for the second half of February, 'J' is for the first half of May ('I' is not used), and so on with 'Y' denoting the second half of December. The first half is always the 1st to the 15th of the month, regardless of the number of days in the second 'half'. The second letter and the number indicate the order of discovery within that half-month. The first asteroid discovered in the second half of February 1950, for example, would be provisionally designated 1950 DA. Since more than 25 objects (again, 'I' is not used) might be detected within a half-month, a number is also appended which indicates the number of times that the letters have cycled through. Thus, the 28th asteroid discovered in the second half of March 1950 would be 1950 FC1 (where F=March and C1 denotes one full 25-day cycle plus 3 (A, B, C), while 2002 NT7 was observed in the first half of July 2002 and was the $T = 19 + (7 \times 25)$ or the 194th object discovered during that time.

On 14 June 2002, the 100-meter 2002 MN object passed by the Earth at a distance of only 120,000 km, one-third of the distance to the Moon! Not only is the distance surprisingly short, but also the fact that the object was detected 'after the fact', three days later. Its energy was equivalent to 180 million tons of TNT. Even with present techniques, unfortunately, 2002 MN could not be

detected sooner! Asteroid 2002 NT7 was even more frightening with a size of 2 km, only 5 times smaller and about 100 times lighter than the dinosaur killer! It was detected on 9 July 2002. The energy liberated at impact would represent several billion tons of TNT. The object has been classified as 'positive' on the Palermo scale, and was the first object to have a positive sign on that scale. Its orbit would cross that of the Earth in 2019 according to computations derived after six consecutive days of observation. Within this uncertainty margin, the chance of an impact would have been 1 in 200,000. Fortunately, following more refined observations, since August 2002 the object has been scaled down and is no longer considered dangerous.

Box 3.3 The Palermo and Torino scales

The Palermo Technical Impact Hazard Scale categorizes and prioritizes potential impact risks spanning a wide range of NEO impact dates, energies and probabilities, quantifying in more detail the level of concern. Its name recognizes the historical pioneering contribution of the Palermo observatory to the first asteroid observations.

The scale is logarithmic (both positive and negative values are allowed) and continuous. It incorporates the time between the current epoch and the predicted potential impact, as well as the object's predicted energy and compares the likelihood of occurrence of the hazard with the average random risk – or background risk – posed by objects of the same size or larger over the years until the predicted date of impact. A value of minus 2 indicates that the predicted event is only 1% as likely as the random background hazard; a value of 0 indicates that the single event is just as threatening as the background hazard; and a value of +2 indicates an event that is 100 times more likely than the background impact. The scale has integer values from 0 to 10.

The Torino scale, so called because it was adopted in that city in 1999, is designed to communicate to the public in a more qualitative form the risk associated with a NEO. Objects are first prioritized according to their respective value in the Palermo scale in order to assess the degree to which they should receive additional attention (i.e. observations and analysis). Colors are associated to these numbers, ranging from white (zero hazard) to red (certain collisions, 8–10) through green (normal,1), yellow (meriting attention by astronomers, 2–4) and brown (threatening, 5–7).

Asteroid 2004 MN4 is better known under the name Apophis, the Egyptian god who threatened the cosmos and attacked the boat of the Sun. The potential impactor of 270 meters would come within about 32,000 km of the Earth in 2029, closer to Earth than the geostationary orbit at 36,000 km altitude. It has been estimated that if its trajectory crosses a small region of space called a

'keyhole' of just a few 600 meters, where Earth's gravity would perturb the asteroid's trajectory, it would definitely encounter the Earth again on 13 April 2036. The probability of passing through the keyhole was about 1 in 45,000 as of February 2007. This places Apophis at the level of minus 2.52 on the Palermo scale or 0 to 1 on the Torino scale, therefore not too much of a concern. This estimation, however, depends on the accuracy of the asteroid's orbital period (30 seconds over 425.125 days) which is extremely arduous to evaluate because the asteroid is one of those that reside inside the Earth orbit, most of the time in full sunlight. The damages on Earth will probably not be global if the size of Apophis is confirmed, but could create havoc on large parts of the globe. More accurate radar observations will be made in January 2013 when the asteroid reappears, hopefully narrowing down more accurately its trajectory and the dates of its future passes. A deflection effort for Apophis (see below) prior to the 2029 keyhole passage would require more than four orders-of-magnitude less momentum transfer than after 2029, and good tracking data during the 2012–2013 apparition is particularly critical for refining impact probabilities and deciding whether a deflection action is required before the 2029 close approach.

Table 3.3 lists the most dangerous potential NEO collisions for the next 100 years. 2004 VD17, a 580-meter object, represents the most pressing danger. It may impact the Earth on 4 May, 2102, liberating an energy equivalent to 10,000 megatons of TNT. When it was first observed, 2004 VD17 was classified as 'green' on the Torino scale, and is now 'yellow' after better observations have been made, meriting careful attention from astronomers. This list will never close because more precise data and observations are continuously appearing. Most likely, the proximity of these events simply reflects the fact that we are observing the sky more accurately now than in the past (we are not aware of the objects that most probably came even closer to us at earlier times). Indeed, the incursions of these celestial visitors are becoming more visible, and for most of

Table 3.3 Most dangerous potential NEO collisions forecasted for the next 100 years. Impact probabilities are cumulated over the number of events for the time span of close conjunctions. c_T is the Torino scale of the risk associated with the object (Chesley et al. [23])

NEO name	Time span	Events	c_T	d (km)
2004 VD17	2091–2104	5	2	0.580
2004 MN4	2029–2055	9	1	0.270
1997 XR2	2101–2101	2	1	0.230
1994 WR12	2054–2102	134	0	0.110
1979 XB	2056–2101	3	0	0.685
2000 SG344	2068–2101	68	0	0.040
2000 QS7	2053–2053	2	0	0.420
1998 HJ3	2100–2104	3	0	0.694
2004 XK3	2029–2104	66	0	0.040
1994 GK	2051–2071	7	0	0.050

them, if not all, we should now attempt to forecast their degree of danger as precisely as possible.

3.3.5 Mitigation measures

The principal issue is then to decide what mitigation measures should be planned. One option is to do nothing and just wait for the unavoidable disaster and prepare for it. A more proactive option is to tackle the problem at its source and work on the impactor itself. The sooner we know its trajectory, the higher the chances of success of mitigation. The precise knowledge of the orbit is essential to be sure that any maneuver will not place the potential impactor on an even more dangerous orbit. Two options can be envisaged to get rid of the danger: either destruction or deviation of the impactor. Of these two, the former is just as dangerous as the direct collision of the object itself because it might spread fragments as large as a few hundred meters along unknown trajectories, as was the case of the Shoemaker–Levy comet impacting Jupiter. The damage could be dramatic on Earth in unprepared areas. Edward Teller, father of the American atomic bomb, had proposed a most dangerous variant to this option creating a collision between a small asteroid and a large one, breaking the former into many smaller pieces [24]! NASA's Deep Impact mission has demonstrated that it is possible to shoot projectiles to a comet, pre-figurating future possible mitigation strategies (Figure 3.15). It seems, however, that the deviation option is clearly the best and probably the only method.

3.3.6 Deviation from the dangerous path

Deviation requires, first, a rendezvous with the impactor. This might be difficult to achieve for several of them, especially if they are on retrograde orbits as is the case of Halley's Comet. Giotto's encounter in March 1986 had a velocity relative to the comet of some 69 km/s. At that velocity, the fine dust surrounding the nucleus can destroy the spacecraft or part of it when it crosses the coma, and indeed this is what happened to Giotto whose camera was destroyed by the dust impacts. Hopefully, in a large number of cases, the object can be approached through a soft rendezvous or even a landing, as has already been demonstrated by NASA and the Japanese Space Agency. ESA with its Rosetta probe plans to land by 2013 on Comet Churyumov–Gerasimenko – the names of its two discoverers – to analyse the nucleus *in situ* as well as its dust and gas as the comet is gradually heated when it approaches the Sun (Figure 3.17). These examples prove that the problem of rendezvous with a NEO is not a major difficulty. However, launch windows must be respected, as the asteroid is not exactly 'on call'. This is an important element to take into account while scheduling a mitigation operation.

Several concepts are presently being considered for deflecting an asteroid found to be on a probable impact trajectory with Earth. For example: applying a velocity perturbation parallel to the orbital motion of the object, changing the characteristics of the orbit and its period so that it reaches the Earth earlier or

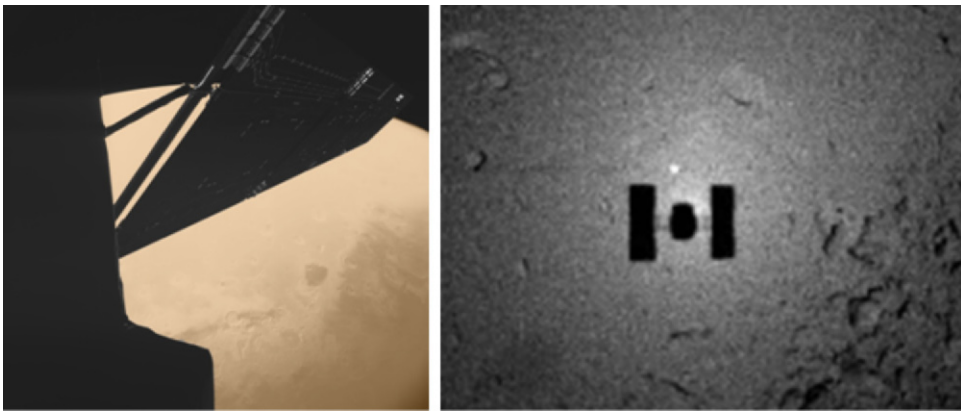


Figure 3.17 Unique images of two interplanetary missions to small bodies. Left: ESA's Rosetta probe imaged by its own Imaging System (CIVA) placed on board the Philae lander just 4 minutes before the spacecraft reached closest approach, 250 km, to Mars on 25 February 2007 during the gravitation assist manoeuvre of the spacecraft around the Red Planet (seen in the background). Right: Shadow of the Japanese Hayabusa spacecraft illuminated by the Sun, projected on the surface of the Itokawa asteroid a few moments before landing from an altitude of 32 meters. (*Credit*: ESA and JAXA.)

later than the forecasted encounter [24]. After completion, some time is necessary to properly evaluate the new characteristics of the orbit, and that time must also be taken into consideration when establishing the schedule of the operation. Orbit perturbations can be grouped into two main classes: kinetic impact and gravitational tolling.

Kinetic perturbations

Kinetic perturbations can be induced by striking the object with large masses, as was done by NASA's Deep Impact mission in the case of Tempel 1. The impact, however, did not induce any measurable velocity change to the 14-km comet in a proportion of 10^{-7}. This, in principle, should be more efficient than exploding the whole object, assuming that the bombardment does not result in the release of large chunks of debris. The conservation of momentum would slightly displace the NEO from its course under the reaction conveyed by the material ejected from the crater in the direction opposite to the impact, which depends on the nature of the asteroid, its physical dimensions and the energy of the projectile. A 200-kg projectile with a velocity of 12 km/s, impacting a 100-meter asteroid with a mass of 1 million tons would perturb its velocity by some 0.6 cm/s, assuming at least an amount of 100 tons of ejecta from the crater for a projectile with an equivalent explosive energy of 1 ton [25].

Another strategy, to use over a long time, would be to attach some kind of rocket nozzle to the NEO to accelerate it. Another concept, called 'mass driver', would be to mine the asteroid and eject the material with velocities larger than

the escape velocity [26]. Operating these strategies for a 1-km object over a decade for an escape velocity of 0.3 km/s would require a total mass of 7,000 tons of ejected material to induce a 0.2-cm/s perturbation. One difficulty is that asteroids are rotating and the mass driver would have to be activated at proper times to change the trajectory and not the spin, making an operation of this system fairly complex. Similarly, radiation from the Sun, or powerful lasers properly focused on the light-absorbing areas of the asteroid, would heat its surface and create vapor or dust plumes that would act as small rocket engines in exactly the same way as comets develop their coma when they approach the Sun and slightly modify their trajectories. However, the use of lasers for de-orbiting large space debris (see below) that are, in fact, much smaller than any of the possible harmful NEOs, is not presently feasible essentially because the masses of the debris are too large for available lasers. Therefore, the technique does not seem to be applicable for large objects. Some ingenious people have even proposed painting the asteroid to change the amount of solar radiation it reflects, thereby altering the forces acting upon it and eventually curbing its course. This is more a fictitious than a serious option because of the amount of paint that would be required.

All these options show that a good knowledge of the properties of the asteroid and its material is required. Solutions are studied in Europe, in particular the Don Quijote project at ESA (see Box 3.4), to analyse the properties of a potential impactor prior to modifying its trajectory.

Box 3.4 Don Quijote

ESA is studying 'Don Quijote', a precursor to a mitigation mission. It uses two satellites launched by the same rocket to deviate a 500-meter asteroid. The first one, 'Sancho', would have a faster velocity and would reach the asteroid first and observe it for several months, also depositing instruments on its surface such as penetrators and seismometers. The second one, 'Hidalgo', would impact the asteroid with a relative velocity of 10 km/s. The impact and its effects will be monitored by 'Sancho'. Analysis of these observations will give a better understanding of the internal structure of the NEOs in view of adjusting a mitigation intervention at a later stage.

Ablation

Ablating the asteroid through the bombardment of particles emitted by nuclear explosions seems to offer, in principle, a more efficient option which furthermore would not depend upon the nature and physical properties of the NEOs. The principle would be to erode part of the asteroid's material through the impact of neutrons from a nuclear explosion above the object over a large area of its surface. This instantaneous blowing off of a superheated shell

of material would impart an impulsive thrust in the direction opposite to the detonation. For an optimal detonation altitude of half the 'radius' of the object, a shell of 20 cm thickness, encompassing about one-third of the asteroid area, might be blown off. It is estimated that deflection velocities of 0.1 cm/s for asteroids of 100 meters, 1 km and 10 km, respectively, require between 0.01–0.1 kilotons, 0.01–0.1 megatons and 0.01–0.1 gigatons of explosive energy [25]. However, the use of nuclear explosives in space is highly problematic, not only because it is an explicit violation of established international law, but also because its effects are highly uncertain. Therefore, the political aspect of that solution is a non-trivial issue. Nevertheless, it remains a last-ditch option in case the additional technologies needed to provide a more acceptable capability are not available.

Gravity tolling

One concept is to attach a robotic tug boat to the NEO and push it out of the Earth's path with the help of an ion engine which operates for a very long time. The thrust would probably be quite small, but when activated for a sufficient amount of time, and sufficiently early, the engine could be strong enough to deflect a NEO up to 800 meters across. In this method it is necessary to have precise knowledge of, first, the physical properties of the object to properly attach the engine and, second, of its orbit to avoid placing it on an even more dangerous course.

The following concept, called 'gravity tractor', is probably the most novel and imaginative of all approaches to overcome this difficulty [27]. Instead of landing on the asteroid, the tractor would hover above it, using the gravitational attraction between the probe and the object as a tow line to gradually change its path. The scheme is insensitive to the surface properties of the NEO and to its spin contrary to the kinetic impact. The ion engine must be actively throttled to control the vertical position of the probe, which is unstable. One important factor is the proper orientation of the nozzles on the two opposite sides of the tug so that their exhausts do not impact the surface of the NEO. It is estimated that a deflection of only 10^{-6} m/s, 20 years before the closest approach of Apophis in 2029, would be sufficient to avoid a later impact. This could be accomplished with a 1-ton tractor exerting only 0.1 newtons of thrust for just a month [27].

Many of these options are still in the realm of fiction. Assuming the availability of current technology, it is only a portion of the potential threat (unknown in size) that can be deflected. A comprehensive protection capability will require additional and substantial technology investments. However, the development of higher performance advanced propulsion would seem to be critical to the eventual availability of a comprehensive planetary protection capability.

3.3.7 Decision making

All efforts undertaken as of now, or still under development, are on a search and observational basis since no world organization has been mandated to evaluate the reality of the threat of a NEO impact and what should be the most appropriate mitigation measure. Everything is mostly theoretical, as there is, as yet, no central command and control system in operation. Probably the most experienced organization in that field is the military because of their missile warning activities. This is therefore mostly in the hands of one or two countries,

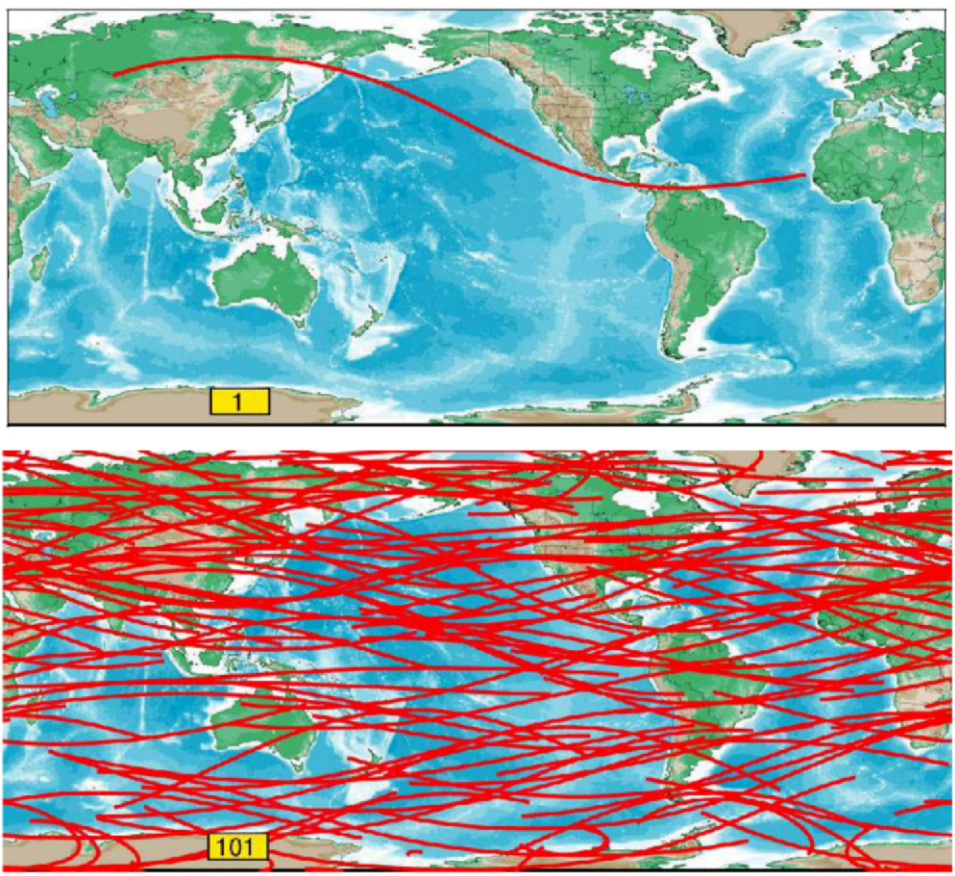


Figure 3.18 Top: Apophis path of risk. The red line defines the narrow corridor within which, if it impacts the Earth on 13 April 2036, Apophis will hit with a probability of 1 in 45,000. Bottom: Set of paths of risk for the 100 NEOs of comparable concern to Apophis, anticipated by the completion of the survey in 2020. Virtually every country in the world will be at some risk, thus illustrating the need for international cooperation in any deflection decision. (*Credit*: R. Schweikart, see reference [20].)

the USA and Russia, but China could also be involved. For example, the US Defense Department is developing a NEO Warning Center to be folded into the US Strategic Command. This is certainly not ideal since, as illustrated on Figure 3.18, all nations are concerned because of the uncertainties of observations and of the estimated impact location [20]. If a NEO is confirmed to be on an Earth-crossing trajectory, the Path of Risk – the most probable line of impact on Earth – will cut through several nations, with the remaining uncertainties possibly increasing the number of those concerned. Even though the uncertainties will naturally shrink with time as more precise measurements become available, the necessity to obtain the best possible information on the trajectory of the NEO as soon as possible is very critical as the earlier the decision is taken to divert it, the better the chances of avoiding an impact. In that respect, it has been proposed to install radiowave emitters attached to the asteroid to follow its displacement with a very high precision. Another concept would be to develop a dedicated GPS network around the asteroid whose position would be determined with respect to a set of distant stars serving as an absolute reference.

No single nation will or should take the decision alone to shoot down or to undertake a deviation mission, especially if they do not possess the proper means. This will never be an easy decision: should it be taken for a specific NEO for which the probability of impact is 1 in 10, or 1 in 100, or 1 in 1,000? The set of nations concerned will have to accept a possible higher risk if the maneuver is either interrupted or only partially successful. The mitigation decision should naturally be agreed by all the nations concerned. In the present context and for some time in the future, several international organizations or agencies will have to be involved. But the long-term and permanent solution goes through the setting up of a dedicated international organization under the aegis of, for example, the United Nations. We address this issue in Chapter 11 and in the general conclusion of the book.

3.3.8 Space debris

Although of no direct cosmic origin, man-made objects are also presenting a threat because of their high speed, which ranges between 15 and 20 km/s (Figure 3.19). Since the launch of Sputnik-1 more than 4,500 additional launches have taken place resulting in a tremendous accumulation of space debris. This requires a global approach to either eliminate them or stop their accumulation [28]. The problem is not with hazards on the ground, although the fall of a major piece of space hardware might hit populated areas and cause casualties (the reader may remember the emotion raised by the de-orbiting of the Russian MIR station in March 2001), but it has to do with the potential hazards to spacecraft either manned or unmanned, as even small debris can damage or destroy satellites if they are in collision.

The estimated number of debris include about 22,000 tractable objects larger than 10 cm in all orbits, of which 2,200 are dead satellites or the last stages of the rocket that put them in orbit. They can measure up to some 20 meters in length and 5 meters in diameter. To these must be added about 650,000 pieces of debris

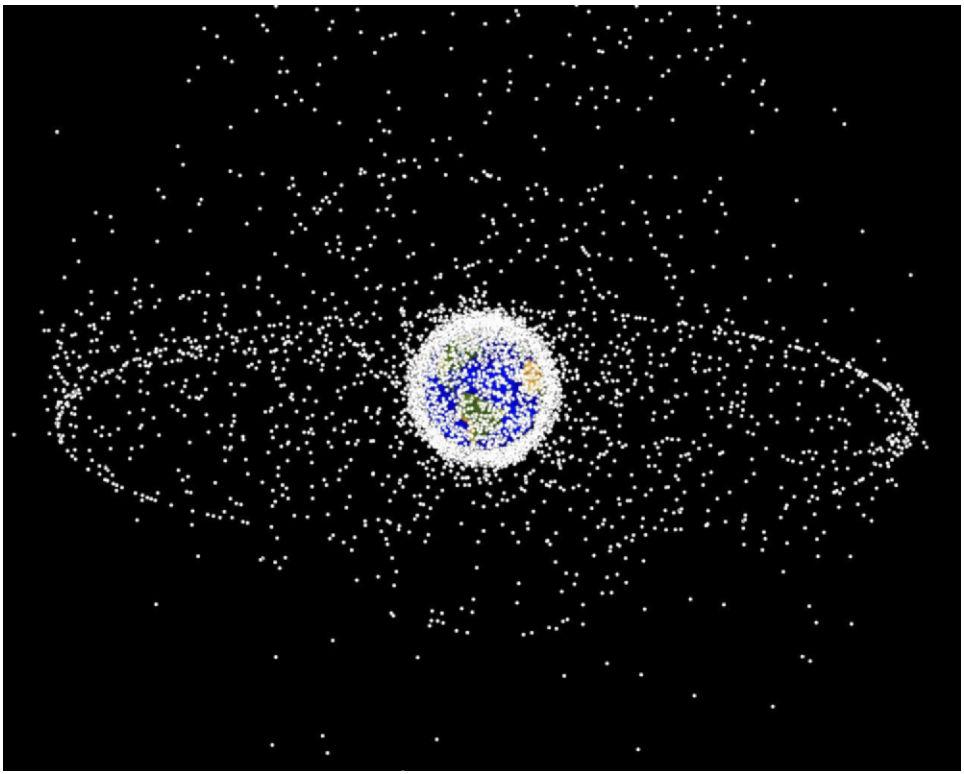


Figure 3.19 Approximately 95% of the objects in this computer-generated image of objects in Earth orbit currently being tracked are orbital debris, i.e. not functional satellites. The dots represent the current location of each item. The ring above the equator corresponds to satellites on the geostationary orbit located at a distance of 35,785 km from the Earth. The images are generated from a distant oblique vantage point to provide a good view of the population. The accumulation over the northern hemisphere is due mostly to debris of Russian origin in high-inclination, high-eccentricity orbits. (*Credit*: NASA.)

of dimension between 1 and 10 cm, and 150 million smaller particles, with a combined mass exceeding 5 million kg. Most of the debris is found at altitudes between 800 and 1,500 km, a region to which a large number of satellites are launched. At these altitudes, the atmosphere is tenuous and cannot exert enough friction to burn them out, contrary to the lower altitude orbits, and debris are there to stay for centuries or thousands of years, maybe even 100,000! At lower altitudes, friction is more efficient and debris burn in the atmosphere after only a few years, depending on their altitude. Debris are systematically tracked by the US Space Surveillance Network (SSN) and cataloged, but only 12,000 of the 22,000 objects mentioned have been cataloged by the SSN. Three accidental collisions between cataloged objects have been documented during

the period from late 1991 to early 2005, which include the January 2005 collision between a 31-year-old US rocket body and a fragment from the third stage of a Chinese CZ-4 launch vehicle that had exploded in March 2000 [29, 30].

The evolution with time of the number of debris can be simulated with the help of models. This is made by both NASA [31] and ESA [32, 33]. According to these models, between altitudes of 200 and 2,000 km, for the period between 1957 and the end of 2200, the population of debris larger than 10 cm will remain approximately constant, with collisions increasing the number of smaller debris and replacing those decaying from atmospheric drag and solar radiation pressure (Figure 3.20). Beyond 2055, collision fragments exceed the population of decaying debris, forcing the total population to increase. About 18 collisions (two-thirds of which will be catastrophic) are expected in the next 200 years. About 60% of all catastrophic collisions would occur between altitudes of 900 and 1,000 km. Within this range of altitudes the number of objects larger than 10 cm will triple in 200 years, leading to an increase of an order of magnitude in collisional probabilities. This is an optimistic evaluation because it does not take into account the increase in the number of future launches and the possibility of accidents like the two that occurred at the beginning of 2007.

On 11 January 2007, China conducted a missile test that knocked into one of its retired weather satellites, creating some 1,100 debris larger than 10 cm and many more of a smaller size distributed between 200 km and more than 4,000 km in altitude. Those higher than 850 km will remain in orbit for at least two centuries. Above 1,400 km, the life time is several thousand years. A little more than a month later, on 19 February, the Breeze-M stage of a Russian Proton rocket orbiting between 15,000 and 400 km, exploded after one year in orbit, creating an equivalent number of debris. The problem is very critical for the International Space Station (ISS) which shares the same 51.5-degree inclination as the Breeze-M stage. Fortunately, the ISS is equipped with bumper shields whose material vaporizes under impact before the main body of the station is hit. In less than six weeks, the total population of debris increased by 20%, evidencing the need for a strict mitigation policy!

As there is no effective way to get rid of debris from their orbit, it is therefore necessary to control their production. The pollution of space is indeed reaching a critical level and most space agencies are now taking preventive measures, such as avoiding explosions in orbit of rocket fuel tanks that are not fully burnt, or de-orbiting satellites to lower altitudes after they have achieved their mission. In 2002, the Inter-Agency Space Debris Coordination Committee, involving the 11 world main space agencies, adopted a consensus set of measures, and in 2007 the United Nations Committee on the Peaceful Use of Outer Space (COPUOS) adopted mitigation guidelines. It is to be feared that these measures will be insufficient to constrain the Earth satellite population and the mitigation guidelines are not legally binding [29]. The removal of existing large objects from orbit might offer a solution to lower the probability of future problems.

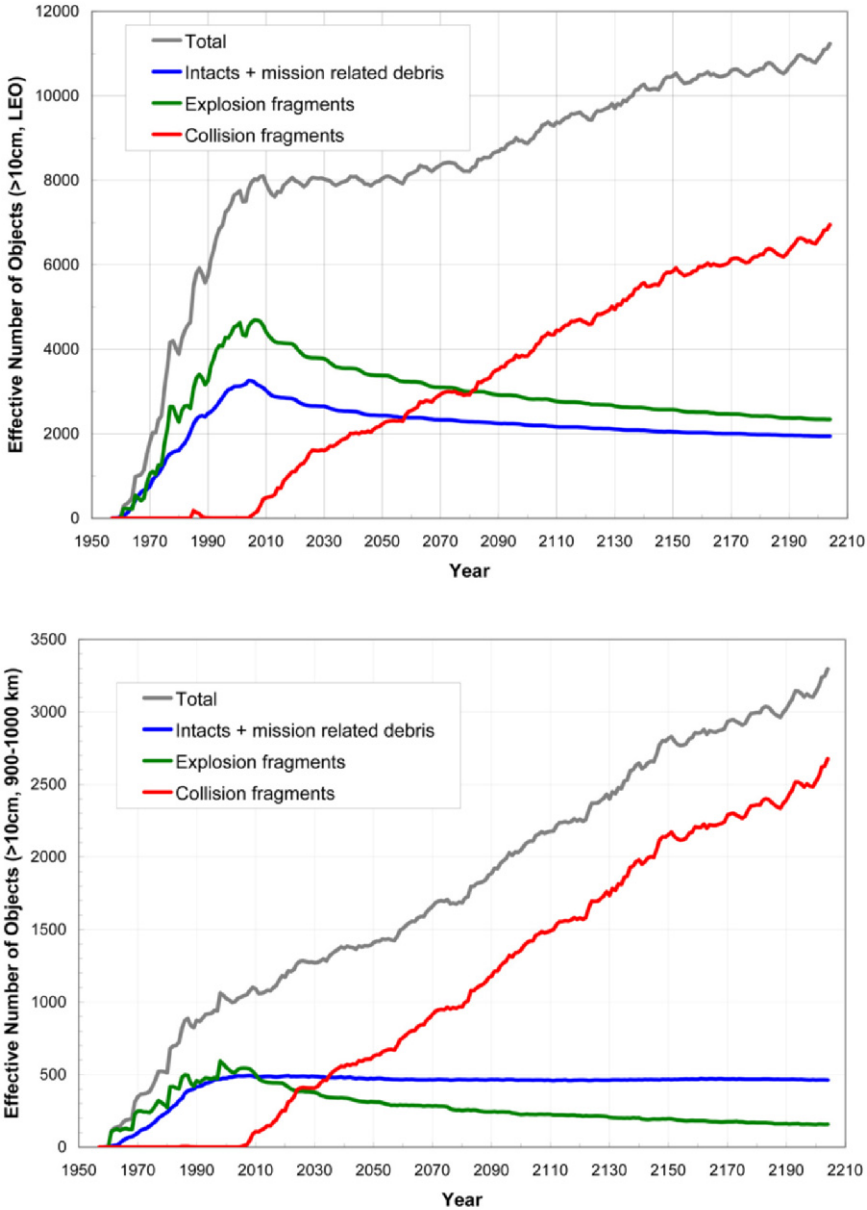


Figure 3.20 Above: NASA simulation of the effective number of 10-cm and larger Low Earth Orbit objects (defined as the fractional time, per orbital period, an object spends between 200 and 2,000 km). 'Intacts' are rocket bodies and spacecraft that have not experienced breakups. Below: Effective number of objects, 10 cm and larger, between altitudes of 900 and 1,000 km. (*Credit*: J.C. Liou et al., see reference [31].)

Unfortunately, no single technique appears to be feasible and economically viable, and research in new technologies is critical in this domain.

In Chapter 9, we mention the possibility of using the Moon for installing a recycling facility of all hardware launched in geostationary orbit. Such a facility would not only recycle the precious materials used in the development of space hardware but also be useful in freeing positions on that orbit. As discussed in Chapter 10, space is an essential asset of all people on Earth in the safeguard of their long-term future, be it for observation, management of resources, navigation, telecommunications, science or manned missions. In that respect, the geostationary orbit is one of the most important. No less critical is the establishment of international regulations prohibiting the testing or use of anti-satellite systems that cannot be naturally cleaned. This appears to be an urgent priority for the future.

3.4 Conclusion

In addition to supernovae, gamma-ray bursts, magnetars, cosmic rays, all being potential destructive agents of the ozone layer, comets, asteroids and space debris, what else could threaten us from the cosmos? Several of these threats have not really worried our ancestors. Paradoxically, it is progress in science and in observation capabilities that have triggered our concerns. Who knows what new instruments and observation systems will contribute to our future knowledge? This cosmic menace may be even more real and stronger than we imagine it today, but fortunately future scientific progress and new technological developments will also raise our understandings, and arm us with the right means of forecast and protection. At that point, we can only be optimistic that these tools will be in operation long before our 100,000-year time limit. This is only one aspect of the problem, however. The other is to have in place the proper management organization to be able to take decisions at the world scale for global security and implementation of the proper mitigation measures.

3.5 Notes and references

[1] Balogh, A. et al., (Eds), 2007, *The Heliosphere through the Solar Activity Cycle,* Springer–Praxis Publishing, p. 286.

[2] The spiral arms rotate at a lower velocity than the Sun around the center of our galaxy which completes a full orbit in 250 million years or so.

[3] Pavlov, A. et al., 2005, 'Passing through a giant molecular cloud: 'Snowball' glaciations produced by interstellar dust', *Geophysical Research Letters* **32,** L03705.

[4] Pavlov, A. et al., 2005, 'Catastrophic ozone loss during passage of the Solar System through an interstellar cloud', *Geophysical Research Letters* **32**, L01815.

[5] Schwarzschild, B., 2002, 'Recent nearby supernovae may have left their marks on Earth', *Physics Today* **55** (5), 19–21.
[6] Benitez, N. et al., 2002, 'Evidence for nearby supernova explosions', *Physical Review Letters* **83**, 081101.
[7] Ellis, J. and Schramm, D., 1995, 'Could a nearby supernova explosion have caused a mass extinction?' *Proceedings of the National Academy of Sciences* **92**, 235–238.
[8] The penetration of solar and galactic cosmic rays has been invoked as causing the formation of clouds in the lower atmosphere of the Earth (see Chapter 10). This is an unresolved issue. If this were to be confirmed, however, the climate might also have been affected by these intense bombardments.
[9] Woosley, S.E. and Bloom, J.S., 2006, 'The supernova-gamma-ray burst connection', *Annual Review of Astronomy and Astrophysics* **44**, 507–556.
[10] Thomas, B.C. et al., 2005, 'Terrestrial ozone depletion due to a Milky Way gamma-ray burst', *Astrophysical Journal Letters* **622**, L153–L156.
[11] Gehrels, T. (Ed.), 1994, *Hazards Due to Comets and Asteroids,* The University of Arizona Press, p. 1300.
[12] Paillou, P. et al., 2003, 'Discovery of a double impact crater in Libya: the astrobleme of Arkenu', *Compte Rendus de l'Académie des Sciences*, doi:10.1016/j.crte.2003.09.008, 1059–1069; and Paillou, P. et al, 2004, 'Eastern Sahara geology from orbital radar: potential analog to Mars', *35th Lunar and Planetary Science Conference*, 15–19 March 2004, League City, Texas; *Lunar and Planetary Science* **XXXV,** 2004LPI....35.1210P.
[13] Varadi, F. et al., 2003, 'Successive refinements in long-term integrations of planetary orbits', *Astrophysical Journal* **592,** 620–630.
[14] Bottke, W.F. et al., 2007, 'An asteroid breakup 160 million years ago as the probable source of the KT impactor', *Nature* **449**, 48–53.
[15] Asphaug, E., 2006, 'Adventures in Near-Earth Object Exploration', *Science* **312**, 1328–1329.
[16] Yano, H.T. et al., 2006, 'Touchdown of the Hayabusa spacecraft at the Muses Sea on Itokawa', *Science* **312,** 1350–1353.
[17] Not less than five dedicated space missions were on course to observe Halley: two Soviets, two Japanese, and Giotto from ESA. Giotto approached the comet the closest to within 600 km of its nucleus.
[18] Keller, H.U. et al., 1987, 'Comet P/Halley's nucleus and its activity', *Astronomy and Astrophysics* **187**, 807–823.
[19] Chapman, C.R. and Morrison, D., 1994, 'Impacts on the Earth by asteroids and comets: assessing the hazard', *Nature* **367**, 33.
[20] Schweickart, R.L., 2007, *Deflecting NEO: A Pending International Challenge,* Presented to the 44th Session of the Scientific and Technical Subcommittee of the UN Committee on Peaceful Uses of Outer Space.
[21] Hill, D.K., 1995, 'Gathering airs schemes for averting asteroid doom', *Science* **268,** 1562–1563.
[22] One optical system is the Lincoln Near-Earth Asteroid Research, LINEAR, a

joint project between the US Air Force, NASA and the Lincoln Laboratory at MIT. It uses two automatic robotized 1-meter telescopes with very fast read-out located at Socorro New Mexico. LINEAR can detect objects of a few hundred meters. Another, Spacewatch, controlled by the University of Arizona, is made of two telescopes in the 1- to 2-meter class. ESA is also operating its Optical Ground Station (OGS) 1-meter telescope in the Canary Islands. The Minor Planet Center (Cambridge, USA) is the international clearing house for observations and orbits of small Solar System bodies, including NEO. It is funded in part by NASA, and its activity is overseen by the International Astronomical Union.

[23] Chesley, S.R. et al., 2002, 'Quantifying the risk posed by potential Earth impacts', *Icarus* **159**, 423–432.

[24] Klinkrad, H. and Grün, E., 2006, 'Modeling of the terrestrial Meteoroid Environment', in *Space Debris – Models and Risk Analysis*, Springer-Praxis Publ., p. 430.

[25] Ahrens, T.J. and Harris, A.W., 1992, 'Deflection and fragmentation of near-Earth asteroids', *Nature* **360**, 429–433.

[26] The velocity that would allow the material to escape and not to fall back on the asteroid from gravitational attraction.

[27] Lu, E.T. and Love, S.G., 2005, 'Gravitational tractor for towing asteroids', *Nature* **438**, 177–178.

[28] Klinkrad, H., 2006, *Space Debris – Models and Risk Analysis*, Springer-Praxis Publ., p. 430.

[29] Wright, D., 2007, 'Space debris', *Physics Today* **60** (10), 35–40.

[30] Liou, J.-C. and Johnson, N.L., 2006, 'Risks in space from orbiting debris', *Science* **311,** 340–341.

[31] Liou, J.C. et al., 2004, 'LEGEND – A three-dimensional LEO-to-GEO debris evolutionary model', *Advances in Space Research* **34** (5), 981–986.

[32] Walker, R.P. et al., 2000, *Science and Technology series, Space Debris 2000*, J. Bendisch (Ed.), Univelt, Inc. Publ. for the American Astronautical Society, pp. xii + 356.

[33] Sdunnus, H. et al., 2004, 'Comparison of debris flux models', *Advances in Space Research* **34**, 1000–1005.

4 Terrestrial Hazards

Bury the dead and feed the living!
Marquis of Pombal

4.1 Introduction

The previous chapter reviewed the hazards of cosmic origins that may threaten living species on Earth. Fortunately, none of them has ever affected human beings or been responsible for a major catastrophe in the past 100,000 years. To the contrary, earthly hazards have clearly left their marks in the history of our civilizations, causing deaths measured by several hundred millions. A distinction must be made between hazards caused by living organisms, at the first rank of which are human beings, and those caused by the 'inert' world or natural hazards which are due to the physical perturbations affecting the solid Earth, the oceans and the atmosphere. In the first category we find wars, which are a fact of humans only, and diseases, both communicable and non-communicable. In the second, we find the seismic-related hazards (volcanoes, earthquakes and tsunamis) and climate-related hazards (storms, floods/landslides and droughts). Figure 4.1 compares the mortality from all earthly catastrophes that occurred in the 20th century – the century for which statistics are more accurate – and places them in perspective. The greatest challenge for all societies will be to predict and manage the disasters that these hazards may provoke in the future.

Wars, with a 20th-century death toll above 200 million – which varies according to different evaluations by several authors and whether national civil fights or massive repressions are considered – represent the biggest danger to humanity. This death toll is more than three times that of epidemics and more than seven times that of famines. Despite this terrible record, wars have been insufficient to stabilize the global population increase. Even though it may have that effect at a local level, the world population has increased by 4.4 billion inhabitants in the course of the 20th century. The war record compares in number with the fatalities resulting from a sub-global asteroid impact (see Table 3.2) whose occurrence however is measured as 1 in several thousand years and causes less than 70,000 deaths per year on average. By comparison, the genuine natural hazards total less than 5% of the number of deaths and are considered to be directly responsible for 3.5 million deaths in the 20th century, not

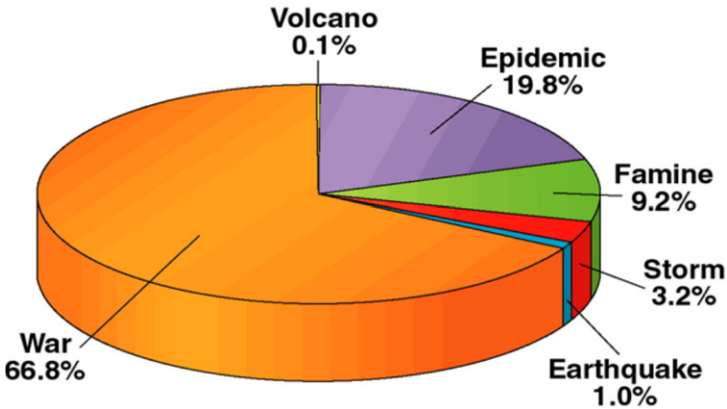


Figure 4.1 Mortality from 20th-century catastrophes. NEOs contribute the same percentage as volcanoes, i.e. 0.1%. (*Credit*: R.C. Chapman.)

considering the secondary causes of related deaths like diseases and famines. According to the United Nations, 75% of the global population is exposed to disaster provoked by droughts, tropical cyclones, earthquakes or floods. According to the 2004 World Disasters Report, published by the International Federation of Red Cross and Red Crescent Societies, 300 million people are affected yearly by natural disasters, conflicts or a combination of both. That was the case in 2004–2005 where, in addition, 300,000 people died just as a consequence of the December 2004 Adaman–Sumatra tsunami. The main course of action clearly seems to lower not only the casualties but also the total number of victims due to both the living world and the natural hazards.

We assume that, in the long term, casualties from wars should definitely decrease. This optimistic view results from the existence of more powerful regulatory control mechanisms, of which the United Nations represent as of now the best approximation that has played a visible and positive role since the end of World War II. Casualties should also decrease from the disappearance of oil as the main energy source, since many conflicts today, and in the recent past, had fight-for-oil as their background. We must also assume that famines and diseases will decrease as a consequence of improved health and hygienic conditions. Similarly, technologies and civil defense policies might contribute to a diminution of all other causes, even though the occurrence of natural disasters cannot be, and will never be, fully controlled – only their consequences, to a certain extent.

Hazards affect the populations of the world differently, depending essentially on their degree of development. This is true for all Earthly hazards including present-day wars both military-regional as well as civilian. Populations are not affected in the same way whether they live in the north or in the south, in the Stone Age or in the Space Age. Figure 4.2 speaks by itself: even though the Americans suffer only 10% of the global burden of diseases, they have the

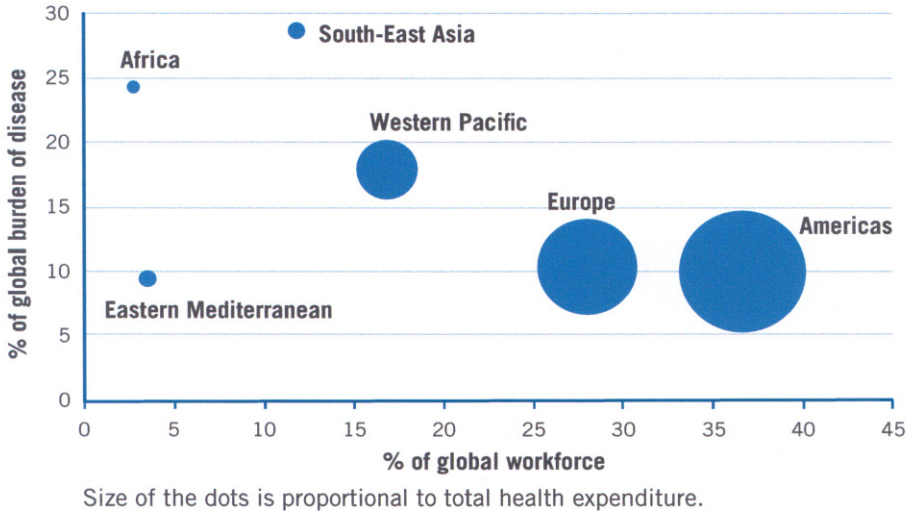


Figure 4.2 Distribution of health workers by level of health expenditure and burden of disease for the various parts of the world. (*Source*: WHO 2006.)

highest percentage of the global workforce and spend more on health than any other part of the world. South-East Asia and Africa represent the other extreme with some 24–30% of the global burden of diseases while having between a few and 10% of the global workforce. Natural disasters happen where there are risk conditions and those conditions are often the result of man-made decisions. Hence, the consequences of disasters are lower in highly developed countries than in less developed countries. Figure 4.3 shows that if the richest countries are the most affected economically by natural disasters because of their more developed infrastructures and the fact that they hold the larger share of the world economy, the poorest have to suffer the biggest burden in inverse proportion of their Gross National Product (GNP). A catastrophe hitting Europe or the USA or Japan today may have more devastating consequences on the world's economy than if it occurred in, say, Africa, even if the death record would be much larger there. Africans may rightfully disagree, but in the future it is expected that their situation will have substantially improved. As said, the death record is not necessarily the best calibre of the magnitude of natural hazards. It just provides an indicative measure of their relative effects on the world population.

4.2 Diseases

Of the 58 million people who died in the world in 2005 of all causes of hazards, diseases do represent about 86% of all deaths. Figure 4.4 gives the respective percentages of the various causes separating communicable and non-

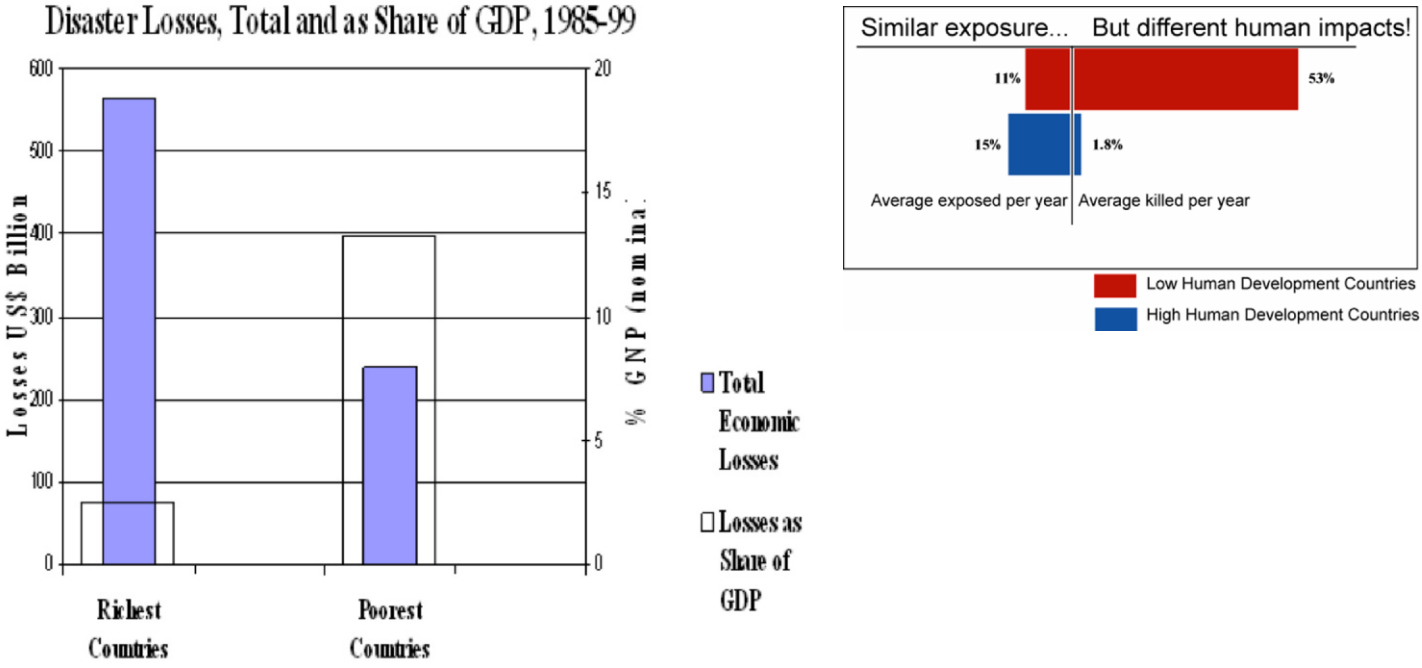


Figure 4.3 Left: Total disaster, economic losses in the world for the period 1985–1999 in billion US dollars and as shares of the GNP. Right: Although the High and Low Human Development Countries are more or less exposed in similar proportions, disaster risk is lower in High Development Countries than in Low Development Countries. Development processes intervene in the translation of physical exposure to hazards into disaster risk explaining why High Development Countries are more exposed than Low Development Countries. (*Source*: United Nations Development Program, Bureau for Crisis Prevention and Recovery.)

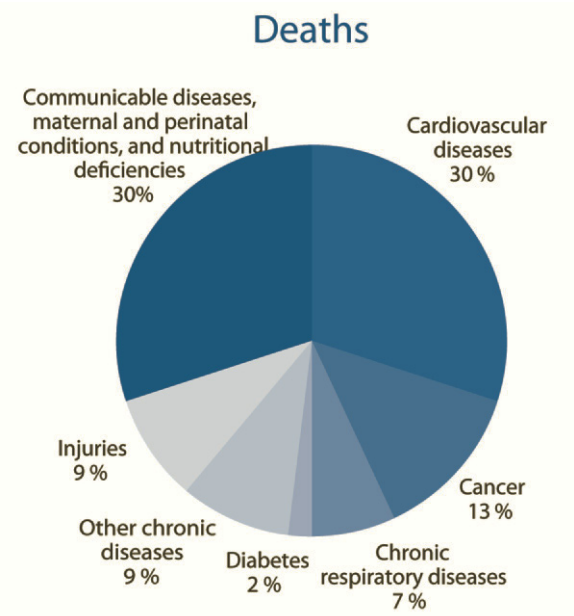


Figure 4.4 The WHO 2005 projections for the main causes of death for the whole world, all ages. (*Source*: WHO 2005.)

communicable diseases. According to the statistics of the World Health Organization (WHO**)**, the latter account for 35 million (of which more than half affect people under 70 years old), which is double the number of deaths from all communicable diseases (including HIV/AIDS, tuberculosis and malaria), maternal and perinatal conditions and nutritional deficiencies combined. A high proportion, 80%, of the non-communicable diseases occur in low- and middle-income countries, where most of the world's population lives, and the rates are higher than in high-income countries. Deaths from non-communicable diseases occur at earlier ages in low- and middle-income countries than in high-income countries. Among the non-communicable diseases, cardiovascular diseases are the leading cause of death, responsible for 30% of all deaths in 2005 – or about 17.5 million people – followed by cancer (7.6 million), and chronic respiratory diseases (4.1 million).

According to projections carried out by the World Health Organization in 2006, during the following 25 years the distribution of deaths in the world will experience a substantial shift from younger to older age groups and from communicable diseases to non-communicable diseases. This is the result of better health care practices, better drugs and advances in research, and better living standards. Large declines in mortality are projected to occur between now and 2030 for the entire principal communicable, maternal, perinatal and nutritional causes, with the exception of HIV/AIDS which are projected to rise from 2.8 million in 2002 to 6.5 million in 2030, assuming that anti-retroviral

drug coverage reaches 80% by 2012. As shown on Figure 4.5 [1], although age-specific death rates for most non-communicable diseases are projected to decline, the aging of the global population will result in significant increases in the total number of deaths caused by non-communicable diseases over the next 30 years, accounting for almost 70% of all deaths in 2030. The four leading causes of death globally in 2030 are projected to be cancer, heart disease and stroke, with HIV/AIDS itself estimated to be responsible for 10% of all deaths. But the death rate of all other communicable diseases will continue to decline. At the same time, the number of people affected by Alzheimer's disease will reach some 2 million.

Communicable diseases are, of course, a main concern in the poorest countries. Whenever possible, vaccination is an obvious mitigation approach that requires political intervention and education of the populations at risk. Basic scientific information, in particular as to what concerns the risks of exposure not only between humans but also with animals, is a very critical measure. Providing the results of scientific research and making vaccines and medicine affordable to the most vulnerable populations would go a long way towards reducing the impact of these diseases. As pandemics know no frontiers, networks for surveillance, information and response also promise to be efficient tools. It is interesting to notice that artificial satellites can now track the geographical extension of diseases, providing data to models forecasting their spread on a global scale.

4.2.1 How old shall we be in 1,000 centuries?

Life expectancy has been regularly increasing as well as life length. In almost every country, the proportion of people aged over 60 years is growing faster than any other age group, as a result of both longer life expectancy and declining fertility rates. Early humans had a much shorter life expectancy than we do today. Few, probably, reached more than 20 to 30 years of age, but by 1900, the average length of life in industrialized nations had doubled relative to this value. In the course of the 20th century, life expectancy in the United States has increased from 50 to 78 years for women and from 45 to 72 years for men, and the death rate per 100,000 inhabitants has decreased by more than a factor of 2 for both categories [2]. Forecasts made by the US Social Security administration put life expectancy in 2050 at 77.5 years for men and 82.9 years for women. In several other countries, according to the WHO 2007 report, life expectancy at birth averaged between both sexes has reached 82 years in Japan, 81 in Australia, France, Iceland and Sweden, and 80 in Canada and Israel – the difference between males and females ranging between 5 and 7 years in favor of females. Overall, it is expected that in 2050, the world population will count more than 2 billion people older than 60 as compared to 600 million in 2007, a number three times that of 50 years before.

Mortality decline has several causes, among which are the growing standard of living, better health and medical care, as well as hygiene and, certainly not negligible, the unique human desire to live longer. Even though the rate of

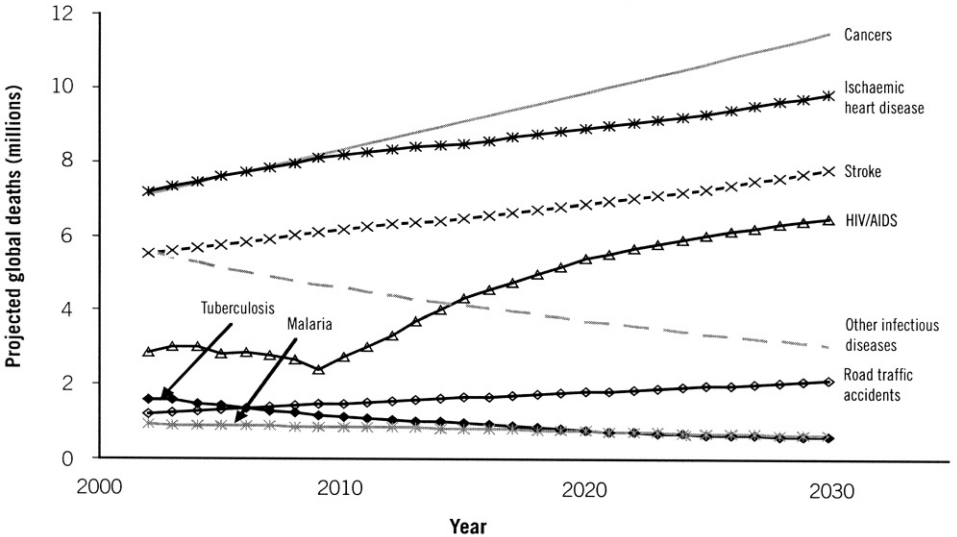


Figure 4.5 Projected global deaths for selected causes of death, 2002–2030. (*Source*: WHO, 2007.)

mortality decline is rather recent and should be extrapolated with caution, over the future 100,000 years we may expect that our planet will be inhabited by a much older population as new technologies and genetic manipulations, as well as anti-aging drugs or practices, may become more accessible and generalized. This population aging can be seen as a success story for public health policies and for socio-economic development, but it also challenges society to adapt, in order to maximize the health and functional capacity of older people as well as their social participation and security. It is then relevant to ask the age to which we will be able to live. In other words, how immortal will we be? Is there a limit to the maximum age of humans, and what is it, if any?

A priori, comparisons between different living species may indeed prove that there is an age limit: mice have a maximum life of 4 years, dogs about 20 years and the oldest human died at more than 120 years [3]. This question is the subject of scientific discussions among a rapidly growing number of specialists as it is very difficult to give a precise answer [4]. Fixing limits to human longevity has in fact little scientific basis, as trends in longevity and in maximal age at death show no sign of approaching a finite limit. However, like machines, the parts of the human body cannot work for ever. But, as is the case for well-designed machines such as, for example, spacecraft, redundancy of parts or of subsystems ensures a longer longevity to the whole system. Switching to the redundant unit in replacement of a defective one may result in a longer life, even though there may be a higher risk of fatal problems when redundancy disappears and vital subsystems are linked in series and are no longer in parallel. Death is then an inescapable destiny. However, as in a well-designed machine, the quality

of its parts may ensure a longer longevity. People who live longer have probably a genetic inheritance (better quality of parts) which slows their aging or protects them from diseases such as cancer or heart attacks that can kill even the hardiest centenarians. The search for genes that might offer a longer longevity and for substances that might slow the 'wear-and-tear' mechanisms is a new branch of medical research. Its promises are still hard to quantify, except that they will most probably yield to ways and practices that will eventually increase longevity, but by how much is difficult to say. Also, the environment in which these parts have to operate is an important factor. Exposing a body to chemical stresses (tobacco, alcohol, air pollution) or mechanical and physical stresses requires both good and efficient repair and maintenance mechanisms proper to the body itself in order to increase its longevity.

The use of the 'machine' analogy to describe the processes and possibly to evaluate the limits of longevity, most probably would lead to overly optimistic estimates as the parts of a living body are themselves more sophisticated than just mechanical or electronic parts, involving in particular a very delicate and complex set of chemical molecular reactions. The human body is made of billions of cells which are aging both individually and collectively. The gradual deterioration of the bioenergetic capacities of these cells (see Box 4.1 [5]) is a determining factor in the aging process, both at cellular and at the organism's level. Numerical computations based on the increase in the frequency with age of the level of this deterioration, indicate that at 126 years of age the totality of cellular mitochondrial DNA would be affected, thereby fixing that limit to human longevity [6]. However, this limit is constantly re-evaluated upwards as more progress is achieved in the understanding of the detailed mechanisms of the aging process, and new measures are tested to slow down that process. Today, specialists evaluate the age limit at around 130 years [7]. That limit may well continue to increase in the course of the next centuries as research continues to progress. Unfortunately, it is impossible to give a firm number. One thing can be said, however: in the future, the Earth's population will include a higher number of very old people than it does at present. This is an important factor to consider when discussing the long-term future to which societies will have to adapt.

4.2.2 How tall shall we be in 1,000 centuries?

Not only do the limits of longevity increase but also the size of humans. One impressive demonstration is found in Europe. In the late 19th century the Netherlands was a land renowned for its short population, but today it has the tallest national average in the world, with young men averaging 183 cm tall. One explanation can be found in the effect of a protein-rich alimentation which has the effect of stimulating the growth hormones. The average size of males and females in various parts of the world range between 170 and 180 cm for males and between 160 and 170 cm for females. This is to be compared to an estimated average size of about 160 cm for our ancestors of 60,000 years ago. The difference is not so great, however, and the recent acceleration in the rate of change would

Box 4.1 The crucial role of mitochondrial degenerescence in the aging process

Mitochondria are the vital element of the energy production of the cell through their utilization of respiratory oxygen. A side product of this energy production is the release, directly in the heart of the mitochondria, of oxygen-free radicals (5% of the oxygen absorbed by the cell's respiratory chain are liberated in the mitochondria in this form) whose high toxicity leads to the deterioration of the mitochondrial DNA which provides, through a continuous and uninterrupted process, the information necessary for the fabrication of the most important units of the respiratory chain. Mitochondrial DNA is totally independent of the DNA in the cell's nucleus, but, contrary to the latter, it has no efficient repair mechanism, is very sensitive to mutations and is 12 times more fragile. With time, this process leads to the irreversible loss of important parts of the mitochondrial genetic material. The result is the gradual deterioration of the bioenergetic capacity of the mitochondria and, ultimately, the death of the cell. The larger the cell, the more important is the risk of mitochondrial deterioration which therefore affects preferentially the large cells such as those of the nervous system, the heart, the muscles, the kidneys, the liver and the endocrinal system. The heart's cells are particularly vulnerable, even for young to middle-aged subjects, as 50% of the total content of mitochondrial DNA of the heart's muscle can be damaged.

Indeed, the energy production of the cells decreases substantially with age, in particular for the nervous system and the muscle cells, which are non-renewable. It has been measured that the loss of mitochondrial DNA with age can reach as much as 85% of the total. It has been shown also that the dysfunction of mitochondrial DNA is directly involved in the apparition of troubles characteristic of aging such as Parkinson's and Alzheimer's diseases. It is also thought that the progressive loss of muscular strength with aging partly finds its cause in the diminution of the mitochondria's bioenergetic capacities.

suggest that we might be much taller than now in 100,000 years. Again, by how much? Today, the tallest man in the world is reported to have been the American Robert Wadlow (l918–1940) with a height of 2.72 meters and a weight of 222 kg before he died at 22 years of age. Most probably his tallness was due to a tumor of the pituitary gland. The tallest living man is reported to be Leonid Stadnik, 36, a former veterinarian living in a small poor village in northwestern Ukraine, and found to be 2.57 meters, beating by more than 20 cm Bao Wishum, a Chinaman of 2.36 meters who previously held the record. Stadnik's growth spurt started at age 14 after a brain operation most likely stimulated his pituitary gland. Bao Wishum apparently owed his size to no endocrinal dysfunction but just to 'natural' causes.

Human height is determined by genetic, alimentation and environmental conditions. It seems that the biomechanical characteristics of mammals, as well as the Earth's gravity, fix limits to the maximum height. In particular the standing position is rather incompatible with a height much larger than 3 meters, although this is difficult to demonstrate rigorously [8]. Excessive tallness is not necessarily an advantage: it can cause various medical problems, including cardiovascular issues, due to the increased load on the heart to supply the body with blood, and issues resulting from the increased time it takes the brain to communicate with the extremities. Mechanical problems are also hampering a serene life.

Reaching higher limits in longevity and tallness is not impossible over the time span of 100,000 years, but is it really an advantage or the sign of a serious degradation of the human body? This consideration would tend to give preference to a situation in which tallness and longevity may continue to increase but would probably not reach respective limits far above 3 meters and a few tens of years beyond the present age record. This, of course, assumes that cosmic as well as natural hazards will not cause a global extinction of life on Earth. Among the latter, volcanic eruptions are probably of greater concern.

4.3 Seismic hazards: the threat of volcanoes

At the end of Chapter 2, we have suggested that some mass life extinction might be attributed to volcanic eruptions. Even though these are local phenomena, their energy and the amount of dust and gases they emit make them some of the most harmful natural disasters at the global scale because of their disturbances on the Earth's climate. Super-eruptions are estimated to occur on average about every 50,000 years, which is about twice the frequency of impacts by asteroids of more than 1 km which are assumed to cause similar effects, making them the most dangerous natural hazards humanity might fear. More than 530 fatal eruptions have been documented in the last 2,000 years with more than 200 such events occurring in the 20th century. The increasing number of fatalities is most likely linked to the increase in global population and not to eruption frequency. Accurate numbers of fatalities are difficult to obtain, in particular (but not only) for the most ancient eruptions, and must be evaluated according to historical records, mentioning such vague terms as 'several' or 'many'. About 275,000 fatalities in total can be attributed to volcanic eruptions in the last 2,000 years, mostly caused by tephra accidents and pyroclastic flows, of which only 2.6% correspond to those where only historical records are available [9]. This is not a big number as none was considered to be major.

4.3.1 Volcanoes and tectonic activity

The permanent displacement of tectonic plates (Figure 4.6) is the cause of earthquakes and volcanic eruptions – and their consequences, such as landslides and tsunamis, are also potentially devastating for the species living nearby

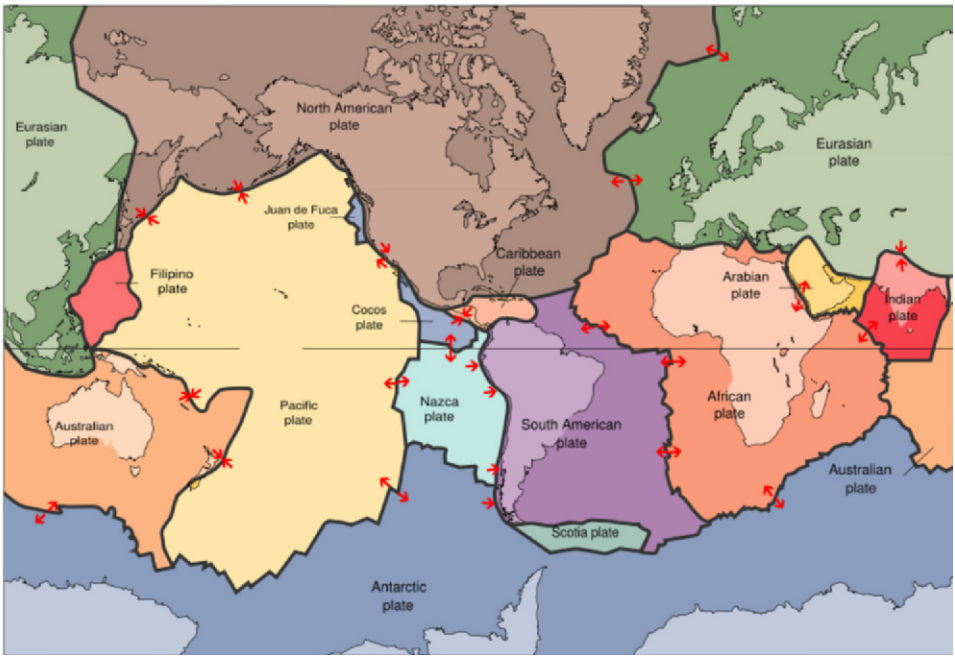


Figure 4.6 The tectonic plates of the Earth. (*Credit*: USGS.)

or further away, depending upon the magnitude of the disaster. The distribution of volcanoes correlate very well with plate tectonics (Figure 4.6), the majority of them being found along diverging plates and rift formation zones, and in subduction zones where one plate disappears underneath a neighbor (Figure 4.7) [10]. In this way, the ocean floor is constantly renewed and new mountains, like the Himalayas or the Alps, are erected.

Volcanic eruptions, however, are not necessarily all associated with the activity of tectonic plates. They are a common feature on several planets in the Solar System or their moons, which are known to have never had, or no longer have, active tectonic plates. Mars, Venus, Io and Titan show evidence of some kind of volcanism. The highest volcano in the Solar System is Olympus Mons on Mars. It culminates at 27 km and has a diameter of more than 200 km at the bottom. Volcanoes may well also be common on extra solar planets; however it is not yet possible to observationally prove this fact.

More than 80% of the Earth's surface is of volcanic origin; at least 1,500 active volcanoes have been identified around the world, and there are probably many more underneath the oceans. That number is increasing regularly as more are continuously discovered. Of the world's active volcanoes, more than half are found around the perimeter of the Pacific, and about one-tenth in the Mediterranean area, Africa and Asia Minor. Per country, Indonesia has by far the largest number of volcanoes, followed by Japan and the United States. The

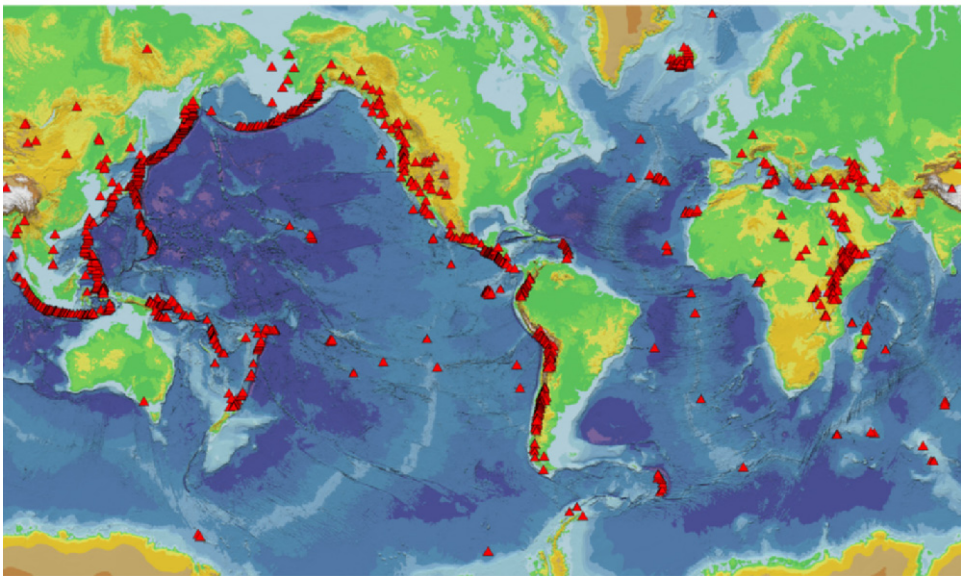


Figure 4.7 Distribution of volcanoes on Earth. (*Credit*: Smithsonian Institution, Global Volcanism Program, Digital Information Series, GVP-3.)

biggest volcano on Earth is probably Mauna Loa, in Hawaii. It rises to 4,300 meters above sea level or about 10,000 meters above the seafloor. Volcanoes grow in altitude because lava or ash accumulates, adding layers and height. Mt Etna on the island of Sicily in Italy is the highest most active volcano in Europe. With an age of 350,000 years it is probably the oldest volcano on Earth, as most active volcanoes are less than 100,000 years old.

Eruptions are the end of a long process. At about 100 km below the Earth's surface in the lithosphere, the temperature of the rocks is near their melting point, releasing bubbles of trapped gases. These bubbles are the main dynamic elements that cause eruptions. A result of that sudden outgassing is to force the magma to rise through the dense layers towards the surface, using cracks and conduits or fractures between tectonic plates. Grossly speaking, there exist two main types of volcanoes: explosive volcanoes, generally concentrated at subduction zones and continental hotspot areas, and more effusive basaltic volcanoes which are common at mid-ocean rifts and oceanic hotspots (Figure 4.8) [11].

Along subduction zones, friction between the plates and partial melting of the rocks generate a very explosive volcanism with gigantic eruptions such as the very spectacular Pinatubo eruption in 1991. This type of eruption is also observed along the west coast of northern and southern America and in the area of the Indonesian, Philippines and Japanese arcs. Explosive volcanoes produce mostly ashes; their eruptions are called 'pyroclastic'. As the gas builds up behind the solidifying magma, the pressure may become high enough to 'blow the top off',

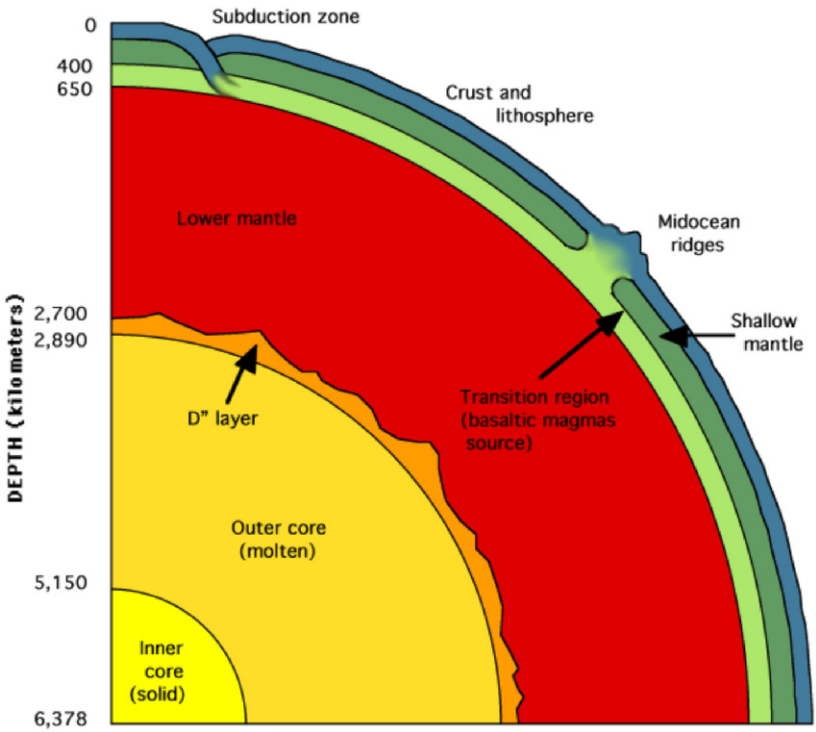


Figure 4.8 Cut through the interior of the Earth schematizing the mechanisms of volcano formation. (*Source*: USGS.)

as was the case for Mt St Helens in 1980. During the eruption, gases, dust, hot ashes and rock shoot up through the opening to altitudes as high as tens of kilometers, with ascending velocities of some 200 km/h and temperatures of 1200°C. These gases are essentially made of water, containing chlorides, carbonates and sulfates as well as carbon dioxide, methane and ammonia. Sulfur dioxide, hydrogen chloride and hydrogen fluoride are also emitted. They are strong poisons and may cause severe problems as discussed below.

Effusive volcanoes produce mostly lava. The lava is just magma that has reached the surface after having lost most of its gases. Icelandic volcanoes pertain to that category. One of the largest eruptions of this type is that referred to as the Deccan Traps in India, which occurred 65 million years ago in close simultaneity with the dinosaur's killer asteroid impact (Chapter 2). Because of the magma's different compositions, there exist different types of lava, with temperatures ranging from 400°C to 1,200°C, and structures ranging from fluid and fast-moving, made of dark basalt (as in the Hawaiian islands, richer in iron and magnesium but silica-poor), to slower and more viscous silica-rich or andesite, so-called because they are typical of the chain of volcanoes found in the Andes in South America. Their velocities are a few meters per second.

Surprisingly, volcanoes are also found in the middle of plates. One common theory explains that type of volcanism by the presence of hotspots or giant plumes of magma that cross through the lithosphere, finding their source from the mantle underneath and melting their way through. The higher temperature, plus probably a thinner crust, may explain the existence of eruptions in such areas that should be expected to be rather seismically quiet. This provides an explanation to the presence of chains of volcanoes of increasing ages, as the motions of the plates gradually cross the hot spots, as is observed in Hawaii, the Tuamotu Archipelago and the Austral Islands. This theory has been challenged by seismologists who question the existence of hotspots and plumes, because they failed to detect them, in particular underneath Iceland. Instead of finding a hot spot there, a broad reservoir of molten rock was detected 400 km down. It is now admitted, however, that hotspots are generally associated with hot and buoyant upwelling, while weaker or lower buoyancy flux hotspots, characterized by lower excess temperatures than the stronger ones (such as Hawaii), can explain the presence of the Azores and Iceland volcanoes [12].

Recently, a new type of volcano has been found far from subduction zones in unexpected places where the bending and flexing of plates opens cracks and micro-fissures such as off the coast of Japan, hundreds of kilometers from where the Pacific plate dips below Japan [13]. These cracks once opened would let the magma go through, creating small underwater volcanoes, also referred to as seamounts. This lends support to the concept that volcanism may appear anywhere on Earth, not necessarily manifesting itself through very energetic eruptions, if we take the Japanese seamounts as model or, closer to us, even the Massif Central in France.

4.3.2 The destructive power of volcanoes

Table 4.1 identifies some of the most characteristic historical volcanoes together with their main properties, in particular their Volcanic Explosivity Index (VEI) which gives an indication of their strength (see Box 4.2). The largest known eruption of the last millennium, the Tambora (1815) in Indonesia, had an estimated VEI of 7, equivalent to 100 km^3 of ashes and debris or tephra. It has been estimated that the total thermal energies of the 1815 Tambora and the 1883 Krakatoa eruptions were equivalent to about 250 Megatons. The Toba eruption of 74,000 years ago in Sumatra (Figure 4.9), most likely the biggest on record, with a VEI of 8, produced $\sim$2,800 cubic kilometers of tephra, more than 2,000 times the amount generated by Mt St Helens in 1980!

About 500 million people presently live close to active volcanoes. Furthermore, the population is growing and the tendency is to settle cities and villages nearby the slopes of these dangerous mountains because the soil there is extremely fertile. In the present world, volcanoes represent a real threat to human life and property. The hazards are of two kinds: (1) local and of short-time duration for the inhabitants close to the eruption, and (2) global and long-lasting as the climate is severely disturbed for several years or even centuries, possibly threatening life on Earth for the strongest.

Table 4.1 Some historical eruptions are listed here together with their main characteristics and the corresponding value of their Volcanic Explosivity Index (VEI), see Box 4.2. (*Source*: Smithsonian Institution, Global Volcanism Program)

VEI	Description	Plume height	Volume of tephra	Classification	How often (duration of continuous blast)	Example
0	Non-explosive	<100 m	1,000's m^3	Hawaiian	Daily (<1 h)	Kilauea
1	Gentle	100–1000 m	10,000's m^3	Hawaiian/ Strombolian	Daily (<1 h)	Stromboli
2	Explosive	1–5 km	1,000,000's m^3	Strombolian/ Vulcanian	Monthly (1–6 h)	Galeras, killed 6 scientists in 1992
3	Severe	3–15 km	10,000,000's m^3	Vulcanian	Yearly (1–12 h)	Nevado del Ruiz, 23,000 deaths in 1985
4	Cataclysmic	10–25 km	100,000,000's m^3	Vulcanian/ Plinian	10's of years (6–12 h)	Galunggung, 1982
5	Paroxysmal	>25 km	1 km^3	Plinian	100's of years (>12 h)	St Helens, 1980
6	Colossal	>25 km	10's km^3	Plinian/ Ultra-Plinian	100's of years (>12 h)	Krakatoa, 1883 Pinatubo, 1991
7	Super-colossal	>25 km	100's km^3	Ultra-Plinian	1000's of years (>12 h)	Tambora, 1815
8	Mega-colossal	>25 km	>2,000's km^3	Ultra-Plinian	10,000's of years (>12 h)	Toba, Sumatra 74,000 yr ago

Even very small VEI eruptions do present a danger, since their lava flows are almost constantly invading what they find on their way, while creating tens of square kilometers of new land. However, the most devastating eruptions are those caused by the highly explosive volcanoes through lateral blasts, lava and hot ash flows, mudslides and landslides, avalanches and floods. Their dust

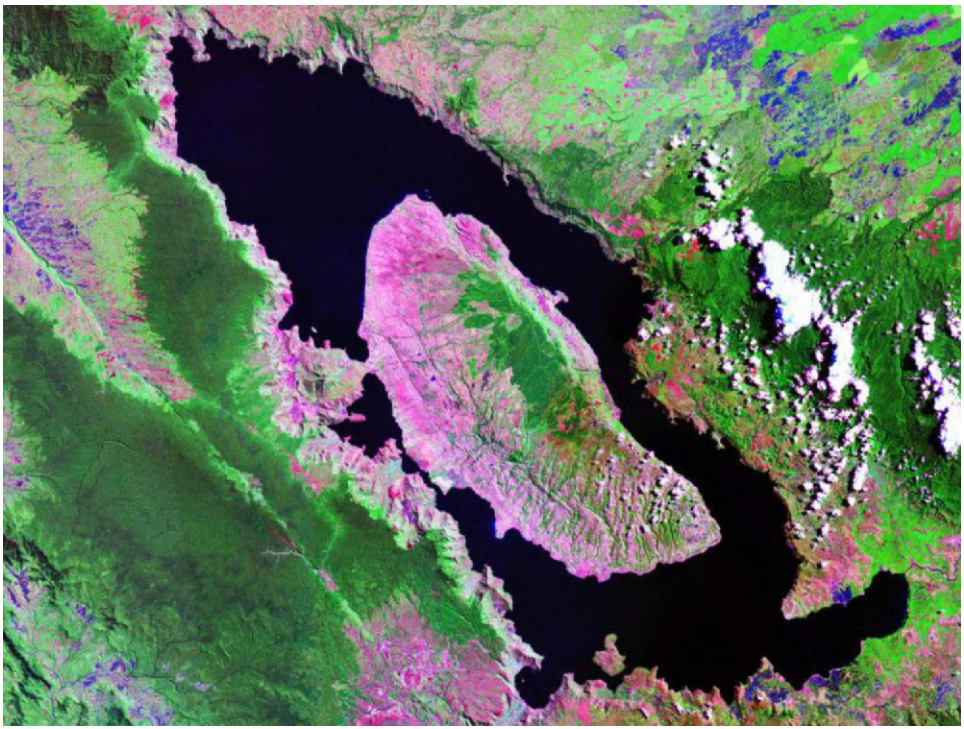


Figure 4.9 The Toba Lake is the largest volcanic lake in the world. It is 100 km long and 30 km wide, and 505 meters at its deepest point. It is located in the middle of the northern part of the Indonesian island of Sumatra with a surface elevation of about 900 meters. Green corresponds to vegetation covered areas, while purple corresponds to arid areas. (*Source*: NASA-Landsat.)

emissions in the atmosphere can also damage the engines of high-flying jets. They can trigger tsunamis or knock down entire forests. The 1902 eruption in La Martinique – the 20th-century's most lethal eruption – remains as a historical example. The lava flows from Mont Pelée were so highly viscous that they may have blocked the throat of the volcano. Pressure build-up from the gases blew out the top in what is the most explosive kind of eruption. In the nearby city of Saint-Pierre, all the 30,000 inhabitants but two died. This dramatic record could have been considerably smaller if it were not for the very poor political management and usage of the early warning indices that were announcing that an eruption was going to happen.

If direct casualties reach rarely above a few thousands – lava flows are moving rather slowly, so that one might usually escape – the resulting acid rains, toxic and greenhouse gases plus famines can account for several tens of thousands. The gas emissions from volcanoes are not very pleasant: sulfur dioxide, hydrogen sulfide, and carbon dioxide can severely damage the health of the surrounding populations, eventually killing them. That occurred twice in Lake Nyos in

Cameroon (a lake crater which formed from a hydro volcanic eruption 400 years ago) in the mid-1980s, as reported in Chapter 2, and where thousands of people and animals died. The Tambora eruption directly killed about 12,000 inhabitants, and it is estimated that a further 90,000 died of starvation, diseases and poisoning. The gigantic 40-meter-high wave of the tsunami generated by the Krakatoa eruption killed more people than the eruption itself. Fortunately, scientific work and the study of past eruptions have helped to save a large number of Filipinos from an early death when the Pinatubo erupted in 1991: only a few hundred people died while tens of thousands were under direct threat. However, this partial success must be tempered by the indirect consequences of living in a devastated country where water is transformed into mud created by loose ash remaining from the eruption.

Box 4.2 Measuring the strength of volcanoes: the Volcanic Explosivity Index

The Volcanic Explosivity Index is based on the degree of fragmentation of the volcanic products released by the eruption or tephra. The greater the explosivity, the greater is the fragmentation of the tephra deposits. Eruptions differ also in the amounts of sulfur-rich gases that form stratospheric aerosols whose climatic effects are important (Chapters 5 and 6). Other parameters considered in the establishment of the index are the volume of the eruption, how long it lasted, and the height it reached. The VEI is logarithmic, so that each number in the scale represents a tenfold increase in the amount of magma ejected out of the volcano.

The scale ranges from 1 to 8, with 8 corresponding to the largest eruptions producing an amount of bulk volume of ejected tephra of $\sim$1,000 km^3. The small values of the VEI from 0 to 1 correspond to non-explosive volcanoes that rarely eject ash and pyroclastics. Hawaiian and Icelandic volcanoes correspond to that category. Logically, small eruptions occur more frequently than larger eruptions as it takes longer to build up the pressures needed for the larger ones. Moderately explosive volcanoes have a VEI of 2–5. They produce lava from basalt, sometimes forming Cinder Cones, but also strato-volcanoes such as the Mt St Helens and Mt Fuji. A Plinian eruption, often called the 'throat clear eruption', completely opens up the 'throat' of the volcano and ashes can reach several kilometers in height.

4.3.3 Volcanoes and climate change

What makes volcanic eruptions particularly noxious is not only the nature and the quantity of tephra and gases that they release, but the altitudes that they are able to reach in the atmosphere, passing through the troposphere and the stratosphere up to several tens of kilometers. Their effect can be either cooling through absorption of solar light or warming, resulting from the large quantities

Table 4.2 Atmospheric effects of some major volcanic eruptions. (Adapted from Rampino [11])

Volcano	Date	Stratospheric aerosols (milliontons)	North hemisphere cooling (°C)
Mt St Helens	May 1980	0.3	<0.1
Agung	March–May 1963	20	0.3
El Chichon	March/April 1982	14	0.3
Krakatoa	August 1883	44	0.3
Pinatubo	June 1991	30	0.5
Tambora	April 1815	200	0.8
Laki	June 1783–Feb. 1784	200	1.0?
Toba	~73,500 years ago	2,200 to 4,400	5 to 15

of greenhouse gases such as carbon dioxide and methane. All climatic effects of volcanic eruptions are discussed in Chapter 5.

Dust and gases are thought to be responsible for a large widespread cooling lasting several years after a major eruption. It is only for the past 100 years that a reliable record exists of the optical depths of clouds of volcanic origin. Dust absorbs and scatters the light from the Sun. Gases such as sulfur dioxide also cause strong effects on the environment. This compound reacts with oxygen and water to produce small drops or aerosols of sulfuric acid which form very highly reflecting clouds and bounce off solar light and reduce rainfall. Table 4.2 presents an estimate of the cooling that resulted from some of the most powerful historical eruptions. A few months after an eruption the global mean sea level falls by several millimeters as a consequence of the cooling which may persist for at least a decade because of the large heat capacity of the oceans [14]. It has been estimated that loading the global stratosphere with one trillion metric tonnes of fine dust and sulfate aerosols would produce a worldwide average temperature drop of about 10°C for several months.

The most recent Pinatubo eruption of 1991 offers a good illustration of these effects (Figure 4.10). It occurred when modern and state-of-the-art measurements became available. Computer simulations have shown that the mushroom from that eruption was as large as 500 km in diameter and extended 35 to 40 km above sea level. Within only a few months, it spread as far as the North Pole, producing the largest aerosol cloud of the 20th century. The terrestrial hydrological cycle was slowed considerably, probably as a result of the reduced incoming solar radiation leading to reduced evaporation. Over the northern hemisphere, the continental surface temperature became cooler than normal by 2°C in the summer of 1992 and warmer than normal by up to 3°C in the winters of 1991–1992. Overall, it yielded a decrease of the global temperature of 0.5°C and a drop of some 6 mm in mean sea level within about a year [15]. However this was mostly unnoticed because at the end of the 20th century, global warming was



Figure 4.10 The eruption of Pinatubo on June 1991 has had major effects on the Earth's temperature and on the ozone layer for more than two years.

already in progress. The result of the eruption may have been just to slow down the anthropogenic phenomenon.

Of some concern is the impact of volcanic eruptions on the ozone layer. Hydrochloric acid, which is produced in massive quantities during volcanic eruptions, is a key chemical agent in the destruction of that layer. Fortunately, it seems to remain confined to the troposphere where the ozone concentration is relatively low. Ozone is one of the most powerful oxidants and a major provider of oxygen to form aerosols from sulfur dioxide. Aerosols themselves provide numerous small surfaces that greatly enhance the destruction of stratospheric ozone by industrially produced chlorofluorocarbons. Nearly one-third of the total ozone depletion could be due to volcanic aerosols at about 17 kilometers. Between 15 and 25 km over the Arctic, these clouds could increase springtime ozone loss by as much as 70%. The resulting increase in lethal UV flux would have dramatic effects on life. Indeed, space observations have revealed a 3 to 8% depletion of the ozone layer after Pinatubo (see Figure 10.5).

The major long-term climatic effect of major eruptions would probably be the collapse of agriculture, as the Earth average temperature would fall by 3 to 5°C, and the cycle of seasons severely affected. Famine would spread, infrastructures would break down and social/political unrest would increase [11]. With proper planning however, provided that such major disasters can be anticipated, the storing of enough food reserves might offer a possible mitigation measure.

However, the present world grain storage is covering only a few months of consumption. It has been estimated that, for a major VEI 8 catastrophe, several years of grain reserves would have to be stockpiled together with other essential alimentation needs, along with maintaining the means for distributing them globally all around the planet.

4.3.4 Forecasting eruptions

Early warning and forecasts are crucial for reducing as much as possible the local and short-term hazards of eruptions. The aim is to save populations and property from immediate effects, ensuring that no one 'is in the wrong place at the wrong time'. However, nothing can be done to avoid the global hazards and, in particular, the long-term climatic effects. The precise timing and magnitude of volcanic eruptions are unfortunately difficult to predict. Historical records of previous eruptions are worth analyzing, even though there are no more dangerous volcanoes than those that are apparently quiet. No clear figure unfortunately is evidenced from the largest VEI 8 eruptions which, in principle, would leave easier-to-identify caldera structures. It could be that such events occur every 50,000 years or as frequently as a few thousand years [11]. On longer timescales, over the past few million years, several studies have found that eruptions seem to follow cyclic patterns in the range of 23,000 to 100,000 years, which correspond to the Milankhovitch frequency band. These bursts of volcanic eruptions would seem to correlate with changes in the climate, sea level and glacier advance and retreat. The loading and unloading of magma chambers by fluctuating ice sheets and sea levels is a possible mechanism that would explain the clustering of volcanic eruptions at times of climate change [11].

Precious to exploit is also the relationship between earthquakes (next section) and eruptions that are linked through the displacements of tectonic plates. New measurements, numerical modeling and statistical analyses support the conjecture that a large earthquake can trigger volcanic eruptions over periods of time extending from days to several hundred years, even at very large distances of several hundreds or even thousands of kilometers [16]. One of the best examples is again the Pinatubo that awoke a few hours after a magnitude 7.8 earthquake, 100 km northeast of the volcano in July 1990, nearly one year before the big eruption of June 1991. The difficulty is to improve the statistical significance of earthquake observations and of eruptions over historical periods. The mediocre reliability of historical records may put the method under serious criticism.

It is fairly straightforward to identify the observations and relevant instrumentation necessary to make the best possible forecasts. The first challenge is to select which of the several hundred active volcanoes to monitor closely. The upward motions of the buoyant hot magma are accompanied by a swelling of the upper layers of the crust at about 5 to 10 km below the surface, and by swarms of earthquakes. The awakening of the volcano is usually preceded by a noise or 'hum', which corresponds to relatively long-period oscillations, ranging between

0.2 and 2 seconds, that seem to be generated when the magma chamber is being re-supplied and starts resonating at these periods. The technique of seismic wave tomography (see Box 4.3 and Figure 4.11) and the knowledge of the frequencies of the waves generated by these pre-eruption quakes as they bounce back and forth on the different layers of underground rocks, permit the acquisition of accurate 'pictures' of the conduits through which the lava will eventually escape to outside. The analysis of the waves is becoming the basic tool presently used to forecast eruptions. This is a 'last minute' indication, however, announcing that the magma is already coming up the conduit, but in no circumstances can it be certain that the eruption will eventually occur because all volcanoes are different and in the course of progression of the complicated pre-eruptive process, account must be taken of all the physical properties of the gases, the rocks and their mechanical resistance, the magma and its viscosity.

These tools are always supported by more powerful computers and more accurate *in-situ* observations. Instruments that are able to monitor variations in the shape of the surface in the vicinity of the volcano or of the volcano itself, as well as temperature variations and vibrations of the ground – such as infrared sensors, seismometers, spectrometers and sample analysis of gases collected *in situ* – provide indispensable information on the nature of the lava and on the imminence of the eruption, since gases escape ahead of the ascending magma. For the first time, Japanese scientists and an international team of geophysicists are drilling into the very heart of the Unzen volcano in southwestern Japan that

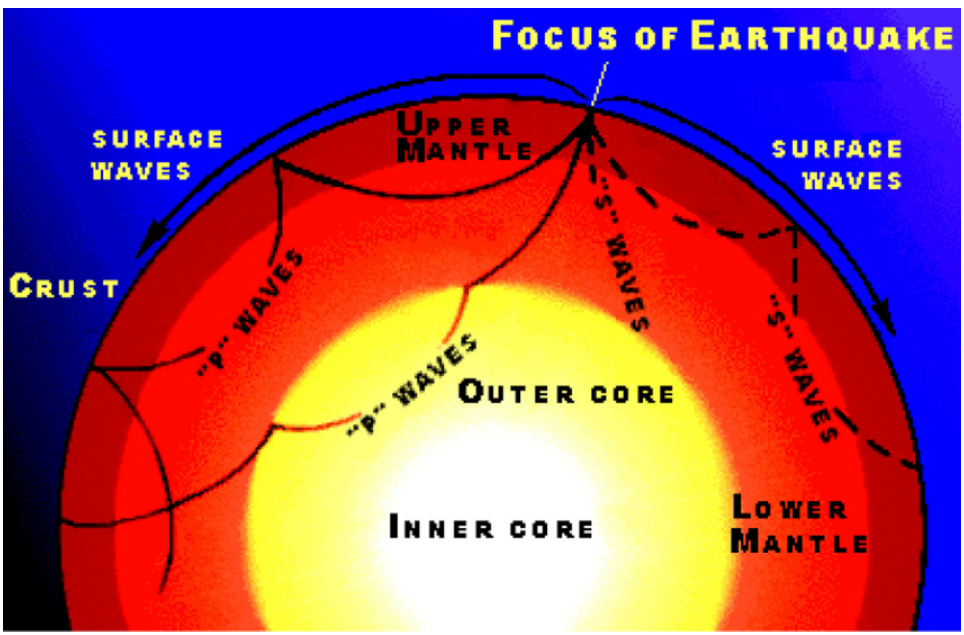


Figure 4.11 The different types of seismic waves. (*Source*: USGS.)

Box 4.3 Seismic waves

The two general types of vibrations produced by earthquakes are surface waves, which travel along the Earth's surface, and body waves, which travel through the Earth (Figure 4.11). Surface waves usually have the strongest vibrations and cause most of the damage. Body waves are of two types: compressional and shear. Both pass through the Earth's interior from the focus to distant points on the surface, but only compressional waves travel through the Earth's molten core.

Compressional waves shake the ground in the direction they are propagating. They travel at great speeds between 1.5 and 8 km per second and ordinarily reach the surface first. For that reason, they are often called 'primary waves' or simply 'P' waves. They travel through rocks, water or liquids.

Shear waves do not travel as rapidly – their velocities are between 60% and 70% of that of P waves – and because they reach the surface later, they are called 'secondary' or 'S' waves. S waves do not reach the deep interior and cannot pass through liquids. Instead of affecting material directly behind or ahead of their line of travel, they displace material at right angles to their path and are therefore also known as 'transverse' waves.

Waves generated by the same quake at different stations allow the determination of the epicenter of the quake. Because the ratio between their average velocities is quite constant, the time delay between the arrival of the P and S waves, allows seismologists to get a reasonably accurate estimate of the distance of the earthquake from the observation stations.

Credit: USGS, RL:http://pubs.usgs.gov/gip/Earthq1/measure.html.

erupted in 1995 with the aim of understanding pre-eruption mechanisms. Ideally, each large volcano should have its own set of instruments, and should be monitored continuously.

Nevertheless, even though some successes have recently been observed, the art of forecasting eruptions is still in its infancy, and volcanoes often defy predictions. Even in the presence of strong signals, the net result might well be a gentle flop! The swelling of the ground may in fact not be due to hot fresh magma coming up, but to the heating and swelling of underground water bodies. That has been the case in the 1970s and 1980s near Naples in front of the island of Ischia, in the caldera of Campi Flegrei that erupted spectacularly 35,000 years ago and is still considered today as one of the most dangerous volcanic areas in Italy. As it is feared that Vesuvius is due for a major eruption in the future, threatening hundreds of thousands of inhabitants of the crowded nearby cities, Italy has established the Vesuvius Observatory, whose main task is to watch the volcano 24 hours a day, 7 days a week and 365 days per year. A first

alarm in 1970, characterized by an increasing number of earthquakes and the rising of the ground by some 60 cm, kept the population very much concerned for about a year, after which the quakes vanished and the ground returned to its former shape.

The success of these efforts is highly dependent upon the level of wealth of the countries under threat, showing again that everyone is not treated the same way and does not share the same chance of surviving. The costs of volcano observatories, as well as the availability of trained specialists, mean that the best predictions will benefit the populations of the wealthiest countries, such as the United States, Japan or Europe. By contrast, Africa hosts some 130 threatening volcanoes that would deserve to be studied, but unfortunately this is not possible as the countries at risk are facing other challenges such as wars or poverty or the volcanoes are just not accessible. This is the case of Cameroon and Congo, which host two of the great threats in Africa, the Mount Cameroon and the Nyiragongo. Both countries have installed observatories on the flanks of their volcanoes, but, unfortunately, civil wars and famines have killed more people in these countries than all the volcanic eruptions of the last century.

However, there is no way that ground-based monitoring can be carried out on all volcanoes around the globe and state-of-the-art technologies, in particular artificial satellites, will become indispensable tools at the service of all vulnerable countries, without having to invest too much money on the ground. These technologies are gradually being introduced in the networks of observational systems. For example, the survey of the variations in the gravity field as carried out with the most up-to-date gravity mapping satellites (Chapter 10), or of the density of the underground magma inside a caldera, as well as of the rising or sinking motions of the ground, are very promising techniques that will certainly be more and more used for the forecast of volcanic activity. If the ground is rising and the density is decreasing, it may mean that the magma is probably becoming gassier, building up the pressure and triggering the explosion. These techniques could potentially give months to years of warning. Ideally, a combination of both ground and space systems would considerably improve the reliability of the forecasts.

Today, super-volcanic eruptions are probably the most dangerous catastrophes we might expect in the future. We can well imagine that a threatening asteroid might be deviated from its dangerous course applying some of the techniques discussed in Chapter 3, but no matter how much we learn from our investigations, there is no solution to preventing a major volcanic eruption to happen. We might be satisfied that we have better knowledge of where it may occur, but in the next 100,000 years the probability of one to occur is not negligible.

4.4 Seismic hazards: the threat of earthquakes

Compared to volcanic eruptions, earthquakes are, generally speaking, local phenomena that mostly affect nearby populations and their infrastructures.

However, when such infrastructures are atomic power plants, the danger can spread on a more global scale. Such a situation did occur on 16 July 2007 when an M6.8 earthquake hit Japan, hopefully with no global consequence due to the precautions taken by the Japanese authorities. Overall, taking into account the uncertainties associated with historical records, nearly 6.5 million casualties have been estimated to be due to earthquakes in the whole world, with China and the Middle East ranking first with more than 3 million for the first, and about 1.5 for the second. On average, about 3,500 people are reported to be victims of earthquakes per year. Their effects on society, however, reach farther than the destruction they cause, because they are sudden and violent, resulting in substantial amount of deaths and casualties, and fear for the populations. Their economic consequences are also potentially disastrous for the whole world in case they hit a major business city as financial places are so tightly interconnected. Their origins – but not necessarily their consequences – are just a fact of nature and not of political decisions, and consequently they draw international solidarity across borders which otherwise would not be permitted politically. That indeed happened after the quake of 2003 in southern Iran, the most violent in the area for more than 2000 years, which caused about 31,000 deaths – when the United States offered to participate in the salvation efforts.

The tsunamis they may generate (see Section 4.5) when they occur near or in the middle of the oceans, might kill large numbers of inhabitants far away. Earthquakes and tsunamis together can destroy or heavily damage vast areas as was the case for the great Lisbon earthquake of 1755 (see Box 4.5 on page 120) or more recently in South-East Asia in 2004. Earthquakes have well-localized origins but this does not mean that they cannot affect the Earth globally. Some may release enough energy that they can cause free oscillations of the entire planet [17] and might change the spin of the Earth, resulting in a decrease in the length of day [18].

Figure 4.12 shows the distribution of past earthquakes on the planet and evidences their connection with tectonic plates. They occur when the plates are in contact with each other inducing brusque movements of large blocks of the crust near the surface. There exist more than 40,000 km of subduction boundaries. Most of the earthquakes take place around the Pacific Ocean but also in the Mediterranean, Turkey, Armenia and central Asia. When the stress increases to beyond that of the brittle lithospheric rock, a rupture appears that may extend over a distance of more than several 100 km, releasing the energy that was stored in the accumulated strain in the form of seismic waves. These waves travel through the rocks, transmitting energy over long distances from the earthquake focus to the Earth's surface and are registered at seismographic stations [19]. The point on the Earth's surface directly above the focus is termed the epicenter. Seismographs help to locate the epicenter and the focus. The intensity of an earthquake is usually given by its magnitude in the Richter scale (Box 4.4) – see also Table 4.3.

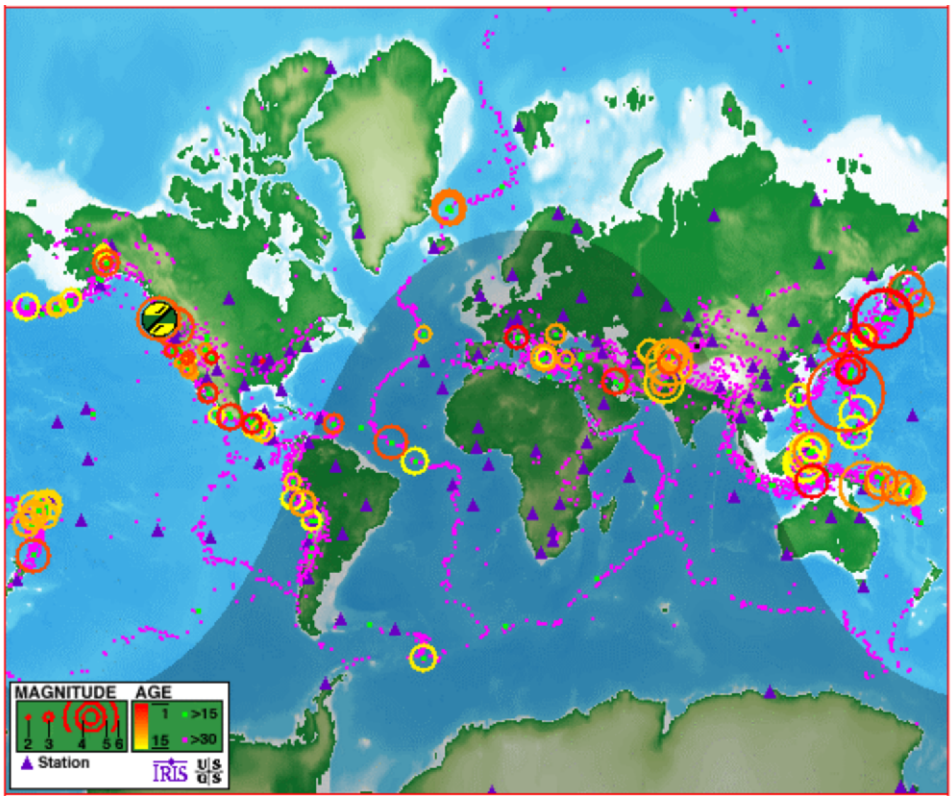


Figure 4.12 Map of major earthquakes as provided by the IRIS program of the US Geological Survey showing their direct connections with tectonic plates. (*Source*: US Geological Survey.)

Table 4.3 Magnitude and frequencies of occurrence of earthquakes

Magnitude in Richter scale	Nature	Frequency	Damages
<2.0	Microseism	8000/day	Not felt
2.0–3.9	Minor	50,000/year	Detected by seismographs; little damage
4.0–4.9	Light	6000/year	Significant damage
5.0–5.9	Moderate	800/year	Major damage to poor constructions
6.0–6.9	Strong	120/year	Destruction within 180 km from epicenter
7.0–7.9	Major	18/year	Severe damage over larger area
8.0–8.9	Great/Important	1/year	Serious damage several hundreds of kilometers from epicenter
>9.0	Exceptional	2/100 years	Very severe damage
10.0	Never observed	None	

Box 4.4 The Richter and Mercali's scales

The Richter scale allows a comparison between the energy liberated at the earthquake's focus estimated from the amplitude of the ground motions measured by seismographs at 100 km from the epicenter. The scale, invented by the American seismologist Charles Richter in 1935, is logarithmic: an increase in magnitude by 1 unit corresponds to a 10 times increase in amplitude of the seismic waves and to a 31 times increase in the energy liberated. An earthquake of magnitude 7 releases 2.10×10^{15} Joules of energy, equivalent to 0.5 Megatons of TNT [20].

Table 4.3 gives the frequency of occurrence and the damages corresponding to each magnitude. A quake of magnitude 2 is the smallest quake normally felt by people. Earthquakes with a Richter value of 6 or more are commonly considered major, and great earthquakes have a magnitude of 8 or more. Earthquakes of large magnitude do not necessarily cause the most intense surface effects because these depend to a large extent on local surface and subsurface geologic conditions.

The modified Mercali scale (seldom used) expresses the strength of an earthquake through its effects. It is not a mathematical scale. It covers the range from I ('Not felt except by a very few under especially favorable conditions') to XII ('Total damage, objects thrown upward into the air'). The evaluation of the intensity can be made only after eyewitness reports and the results of field investigations have been studied and interpreted. The damage from the San Francisco earthquake reached a maximum intensity of XI. (*Source*: USGS.)

The rupture of any one contiguous segment of 800 km or more in length and a few hundred kilometers wide with fault displacements of tens of meters can produce an earthquake similar in magnitude to the Sumatra–Andaman event of December 2004 [21]. Eleven out of the 12 greatest earthquakes ($M \leq 8.5$) of the past 100 years occurred along subduction fault planes. Thousands of earthquakes take place all over the world every day, but few release enough energy to do serious damage. This is an illustration of the Gutenberg–Richter law, which states that small earthquakes occur more often than big ones. Every year, on average 18 earthquakes between magnitudes 7.0 and 8.0 take place somewhere in the world, luckily in unpopulated mountains or under the sea (Table 4.3). The most powerful earthquake ever measured, which occurred in Chile in 1960, had a magnitude of 9.5 but earthquakes of this size are fortunately extremely rare. Table 4.4 gives a list of the major earthquakes that occurred in the course of the 20th century, together with their magnitude and death record.

On surface, an earthquake consists of rapid vibrations that usually last no more than a few seconds, although some may last several minutes, as was the case for the Sumatra–Andaman event, which lasted 9 minutes, the longest ever

Table 4.4 Major casualties and major earthquakes of the 20th century. (*Source*: USGS earthquakes hazards program)

Date	Location	Magnitude	Deaths
2004	Sumatra–Andaman	9.3	283,100
1976	Tangshan, China	7.5	255,000
1920	Haiyuan, China	7.8	200,000
1923	Kanto, Japan	7.9	142,800
1948	Ashgabat, Turkmenistan	7.3	110,000
2005	Pakistan	7.6	86,000
1908	Messina	7.2	72,000
1970	Chimbote, Peru	7.9	70,000
1990	Iran	7.4	40,000–50,000
1985	Mexico	8.0	9,500
1906	San Francisco	7.8	3,000
1960	Chile	9.5	3,000–6,000

recorded in history. The first indication is often a sharp thud, signaling the arrival of the compressional waves, followed by shear waves and the 'ground roll' caused by the surface waves (see Box 4.3). Features such as small brooks or alluvial deposits, valleys or moraines can be shifted by as much as 10 meters on both sides of the fault at each seism, and are useful for measuring the intensity of the quake. Major earthquakes are followed by a series of aftershocks at a rate that decreases with time. They are caused by the strain of the major earthquake not being fully released in the main event and by a rearrangement of stresses in the area.

4.4.1 Measuring the power of earthquakes

An earthquake's destructiveness depends on many factors. In addition to the magnitude and the local geologic conditions, these include the focal depth, the distance from the epicenter, and the design of buildings and other structures. The extent of damage also depends on the density of population. The most deadly earthquake in historical times occurred in China in the Shensi Province in 1556, with a record of about 830,000 victims. The Sumatra–Andaman earthquake and its tsunami killed nearly 300,000 people in Asia in 2004. Together with the 7.6-magnitude earthquake in Pakistan of October 2005, close to 400,000 people died in less than 10 months, which was a heavy toll for the start of the new century. In the Chinese city of Tangshan in 1976, 255,000 died – a quarter of the population – not counting the 150,000 who were severely injured and died later. Japan is also on the black list of earthquake-sensitive countries with nearly 150,000 victims counted in 1923 in Kanto. The 1906 San Francisco quake killed some 3,000 inhabitants, directly or indirectly, out of a population of about 400,000 and almost completely destroyed the city: 225,000 inhabitants were left homeless and 28,000 buildings were destroyed. The fires that followed and lasted for three days caused substantially more damage than the earthquake itself [22].

In comparison, the recent event which affected the city in 1989, had only 64 victims and left the city more or less as it was before! The historical record in Europe is the Great Lisbon earthquake (Box 4.5). Unless mitigation measures are taken, the number of victims of earthquakes and tsunamis is bound to increase in the future as a consequence of the growing population density in the vulnerable areas.

Box 4.5 The Great Lisbon earthquake

The Lisbon earthquake on All Saint's Day [23] struck the capital of Portugal (then the fourth largest city of Europe with a population of 275,000) on 1 November 1755 at 9.45 a.m. It is the most deadly earthquake reported in Western Europe in the last 250 years. Its magnitude was evaluated at 8.7. It resulted from the rupture of a fault about 100 km off the coast of Portugal, not yet precisely identified because of the complexity of tectonics in this area. The quake also affected the city of Cadiz in Spain and the northwestern area of Morocco.

Lisbon was built on unsteady ground and alluvial sands and many buildings sank in the slumping soils. Furthermore, on that special catholic well-celebrated day in Portugal, candles were lit in most houses and a general conflagration engulfed the city. The fires plus the subsequent 5-meter-high tsunami waves which 30 minutes later swept and flooded a kilometer inland, added their tribute to an estimated total of more than 60,000 people killed. Those who survived invoked the vengeance of God in punishment of all the poor 'Lisbonete's' sins.

When King José I asked his Chief Minister what to do, the answer was: 'Bury the dead and feed the living'! That genuine leader, better known as the Marquis of Pombal, also demanded that the clergy stopped preaching that the 'end of day' was near, and then developed visionary plans to rebuild the city, introducing new earthquake-proof techniques. It took about 100 years to complete the work. No one knows of course whether the city is totally earthquake proof today. But this is also the situation of many other places like Tokyo, San Francisco and others, and there is no need for a 'dry run' to test!

4.4.2 Earthquake forecasting

As in the case of volcanic eruptions, it will never be possible to control earthquakes but it might be ultimately possible to forecast them. Given the relatively large number of victims, predicting earthquakes is an essential need and will become even more so in the future as the density of population increases. Accurate short-term predictions for a specific earthquake on a particular fault within a particular year rather than at some unidentified time, within the next decades or so, are the goal of today's forecasting objectives in

Box 4.6 The theories of earthquakes [24]

The Elastic Rebound Model states that at a geological fault between two moving plates, stress occurs and deforms the rocks [25]. This occurs in four main steps, starting from the original position, the build-up of strain, the slippage, and the strain release [26]. If the fault creeps, it will produce frequent micro-earthquakes; if it binds together and then slips, it will produce large earthquakes. Stress will then quickly be released; the sides of the fault will become offset and the rocks will rebound to their initial state of stress. The problem is that earthquakes do not produce the large drop in stress required for this model.

The Seismic Gap Model states that strong earthquakes are unlikely in regions where weak earthquakes are common, and the longer the quiescent period between earthquakes, the stronger the earthquake when it finally occurs. The complication is that the boundaries between crustal plates are often fractured into a vast network of minor faults that intersect the major fault lines. When an earthquake relieves the stress in any of these faults, it may pile additional stress on another fault in the network. This contradicts the model because a series of small earthquakes in an area can then increase the probability that a large quake might follow.

order to be in a position to take all necessary protective measures early enough, and minimize potential casualties and destruction. Would it be wise indeed to leave entire cities uninhabited for several years in the expectation of a possible but uncertain catastrophe? Unfortunately, at present, the prediction of earthquakes is still in its infancy. It is indeed very difficult, if not impossible, to predict the timing of an earthquake even though plate tectonics provide the framework to successfully forecast in the long-term those plate subduction segments that are the most vulnerable and deserve careful attention. An earthquake is the outcome of a series of complex processes, including small events that lead to an amplification of micro-scale phenomena which ultimately result in a major catastrophe, making short-term predictions nearly impossible [27].

Modern research in large earthquake forecasting (see Box 4.6) rests first on the analysis of the history of large events in a specific area and of the rate at which strain accumulates in the rock. It includes field, laboratory, and theoretical investigations of quake mechanisms and fault zones. If a fault segment is known to have broken during a large earthquake, the recurrence time and the magnitude of the next one can be estimated based on the size of the fault segment, the analysis of its rupture history, and the strain accumulation [28]. This method works only for well-understood faults, and is less successful for the others, such as in the case of the 1995 Kobe quake in Japan or of the Sumatra–Andaman event which occurred in a subduction zone where only smaller ($M<8$) earthquakes were historically recorded. Some areas, however, like North

America's Cascadian zone, have no historical records. In places where long histories are available, the frequency of earthquakes appears to be highly irregular. Whenever the ground is quiet for a period much longer than the average recurrent time, there is reason for serious concern. This is because the strain to which the rocks are submitted due to motions and dynamic pressure of the plates is accumulating energy, and the next event will very likely be more catastrophic than its predecessor.

For example, there are evident signs from statistical analysis of past events and on evaluations of the accumulated stress, that a major event is overdue in the Himalayas, putting about 50 million inhabitants in the Ganges plain under serious risk. The affected region lies along the fault line where the northward-moving Indian plate plunges beneath the Eurasian plate at a rate of 2 cm per year, pushing up the chain of mountains. The Himalayas have not ruptured as expected for more than 500 years. When that occurs, at some time between now and 1,000 years in the future, it might lead to slips of more than 20 meters [29]. Certainly, many more such slips can be anticipated in the next 100,000 years! The earthquake of October 2005 in Pakistan did not surprise the geologists, but they were unable to predict its exact timing. In 1999, Turkey was victim of two successive earthquakes. The first one in August of M7.4 caused a surface rupture along a 150-km long fault, 17 km deep, and killed 15,000 people in the city of Izmit. The second, in November – somewhat weaker (M7.1) – occurred at the same place and there is every reason to believe that there will be a new catastrophic earthquake in that area within the next 30 years, and it will be closer to Istanbul, putting about one million people at risk. Since the 1906 earthquake in San Francisco, the Bay Area is relatively quiet. But it is likely that this seismic silence is to end sooner or later as the continued movement of the Pacific and North American plates have reloaded the strain in the faults, leading the USGS to predict a 62% chance of a M7 quake in the area before 2032. In contrast, the Sumatra–Andaman quake ruptured a segment that was among the least likely to fracture [21]. As in the case of weather forecasting, the probability of occurrence will never take the place of certainty. For a long time and in the absence of any firmly-based scientific work, predicting earthquakes was like using a frog in a bottle to predict the weather.

According to reports issued by locals from every seismic region of the world, in particular China and Japan, the behavior of animals has been used as a proxy, as animals were observed to become more excited and more aggressive than usual prior to significant events. Fishermen in these countries also noticed better than usual fishing performances before a quake…! There are also those mysterious fireballs and flashes that were reported to have preceded the M7.6 Tangshan quake of 1976 in China. In all of history, only one single earthquake was successfully predicted: the Haicheng, M7.3 event of 1975 which killed (only!) some 2,000 people. In the months before, signs that something was moving underground could be noticed: the levels of water and of the ground raised sufficiently to be measurable, animals were showing their strange behavior, and there was an increasingly large number of small foreshocks that led the

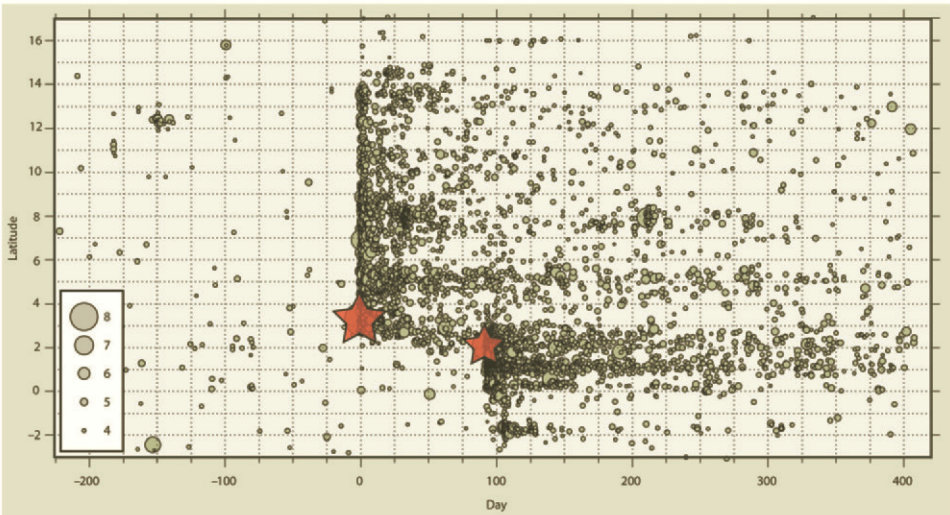


Figure 4.13 Seismic activity before, during and after the great Sumatra–Andaman earthquake (including events occurring beneath the Andaman Sea) as observed with both seismic and GPS methods. Each circle in this space–time diagram refers to an earthquake whose size is represented by the diameter as shown by the scale in the lower left corner. The two largest events in the area, those of 26 December 2004 and 28 March 2005 are represented by two stars and are most probably connected. Such measurements using the most advanced technologies will be of great importance in understanding the earthquake process. (*Source*: USGS.)

authorities to evacuate the city of its one million inhabitants. That, along with the local style of housing construction and the time of the main shock, 7:36 p.m., saved thousands of lives. Was it a real forecasting success or just a coincidence? In fact, the Tangshan event a year later could not be predicted and caused the death of some 255,000 people! More science-oriented observations may therefore be useful in the future.

As empirical as they are, these types of precursor information are nevertheless found to be worthy of consideration by the scientists because they have been observed repeatedly on several occasions, even though it has not been possible to provide an acceptable scientific explanation for their occurrence. Further progress will come from more research into the physics of friction and of the processes leading to rock rupture as well as hydrothermal and magnetic modifications and the escape of gases from the interior of the Earth, before, during and after the quakes.

Earthquake research is clearly one of the areas where complementary ground-based and space-based measurements are required since we need to probe deeply in the interior of our planet and monitor huge parts of its surface (Figure 4.13) [30]. In parallel, modeling, computations and simulations permitted by advances in computer technologies open very promising possibilities. Magnetometers,

strain meters, tilt meters, and water level monitors, make it possible to measure the expansion and the contraction of the ground. Together with laboratory studies, one can therefore get access to heat flow, stress, fluid pressure, and to the mechanical behavior of rocks in the fault zone. Monitoring motions of only a few millimeters per year of the Earth's surface can now be done very accurately with seismic networks. This is extremely useful in assessing the strength of an earthquake in the first few seconds of rupture [31]. The primary seismic waves that radiate from the focus at about 4 to 6 km/s, and are not strong enough to create serious damage, can arrive a few seconds before the more destructive compressive waves hit. Their frequencies might offer the possibility of assessing the magnitude of the entire quake [32]. While a few seconds does not seem to be a tremendous amount of time, it might be long enough to stop trains and factories, and ring alarms, warning that something strong is coming and that people should immediately shelter under their desks or in doorways. But this will not be nearly enough! It supposes that the infrastructures are in place to use this precious information.

New techniques are appearing that bear some hope and promises. Measurements from space to a millimetric precision yield accurate positioning of the Earth's crust, using Global Positioning Systems (GPSs), radar interferometry and space gravimetric systems (Chapter 10), as demonstrated by the measured changes in gravity associated with the Sumatra–Andaman earthquake [33]. Furthermore, space measurements provide methods of connecting several events at very large distances whose coincidence might be unnoticed otherwise. New data may also provide exploitable forecasts in the future. This is the case of peaks of infrared light emissions at about 10 microns of wavelength that have been recorded by satellites above Izmit in Turkey and Bhuj in India in 2001, one to two weeks prior to each respective quake. These emissions may result from mid-infrared luminescence associated with crustal deformation of the rocks through the activation of positive hole-type charge carriers under the high levels of stress prior to the earthquake. They are not the result of heat emission from the ground. Other manifestations of the phenomenon are currently under investigation using geosynchronous weather satellites [34].

Recently, a group of scientists from Taiwan reported variations of the ionospheric Total Electron Content between 1 and 4 days before the M7.7 Chi-Chi earthquake of 20 September 1999 [35]. Such ionospheric disturbances might represent some amplification of ultra-low frequency electromagnetic waves (less than 1 Hz) emanating from magnetic particles inside the rocks as underground stresses build up. In fact, seismic waves emitted during an earthquake have been reported to be detectable at the level of the ionosphere [36]. These observations, if confirmed, may bear some hope that in the near future we might be able to forecast large earthquakes (M>6) several days ahead. Since then, on 29 June 2004 the French space agency, CNES, has launched the DEMETER micro-satellite (Detector for Electromagnetic Emissions Transmitted from Earthquake Regions) which allows the measurement of ionospheric waves and other characteristics of the ionosphere [37]. Unfortunately, even though there is evidence that such

perturbations do exist prior to earthquakes, their signatures are weak and blended with many other sources of highly variable natural and anthropogenic origin – such as cell phones or the Sun, the magnetosphere and auroras – and the usefulness of these signals is yet to be demonstrated through more observations.

4.4.3 Mitigation against earthquakes

With time, earthquakes will most likely become more devastating if not more frequent for the reason that 40% of cities hosting more than 2 million inhabitants are at a distance of less than 200 km from a risky zone. In 2035, the total population of these cities will be around 600 million. Specialists identify China and Iran, in particular the city of Tehran, as places where the number of victims might reach above one million. The areas of Tokyo and California are also likely to be the victims of important economic damage in the near future. Looking at the next thousand centuries, tectonic plates will have remodeled the continents: the Indian plate will have pushed the Himalayas more than 2 km, most likely triggering several hundreds of devastating large quakes, assuming their recurring on a 500-year basis on average.

The first measure is to work on improving the standards of buildings. For example, the M7.1, 17 October 1989 San Francisco quake, the worst since the 1906 event had (only!) 64 victims throughout central California, as compared to the 3,000 in 1906, injured less than 4,000, and left less than 13,000 (as compared to the 225,000 in 1906) homeless. Unfortunately, above a certain limit, no structure can resist a very strong earthquake: a shake of M7.9 in the vicinity of Tokyo, could destroy more than several hundred thousand buildings in the city! Developing the means and civil protection requirements are obviously national policy decisions. It would indeed be prudent for earthquake-prone countries with nuclear power plant capacities like Japan to respect the highest safety measures when building such plants in sensitive areas. When such impacts may affect the whole planet, mitigation and prevention cannot be left to one single government and the need for international cooperation and coordination is obvious here. The Pakistan quake of October 2005, which killed 86,000 people, could have been much less devastating if cooperation had been developed with India, instead of both neighboring countries remaining in ‘splendid seismic insulation’!

4.5 Tsunamis

4.5.1 What are they?

The word ‘tsunami’ can be literally translated from Japanese as ‘harbor wave’ as tsunamis have the ability to penetrate the protected harbors along the coasts of Japan, which are very often hit by them. They are the most dramatic and destructive water waves because they carry gigantic masses of water. Tsunamis get their outermost destructive power when the quake abruptly displaces a large area of the ocean floor several tens of thousands of square kilometers over only a

Table 4.5 Most damaging tsunamis world wide. Statistics before the 20th century are approximate. (Adapted from NOAA)

Deaths	Year	Location	Deaths	Year	Location
230,000	2004	Northern Sumatra	3,800	1746	Lima, Peru
60,000*	1755	Lisbon	3,620	1899	Banda Sea, Indonesia
40,000	1782	Southern China Sea	3,000	1692	Jamaica
36,500	1883	Krakatoa†	3,000	1854	Nankaido, Japan
30,000	1707	Tokaido-Nankaido (J)	3,000	1933	Sanriku, Japan
26,360	1896	Sanriku, Japan	2,243	1674	Banda Sea, Indonesia
25,674	1868	Northern Chile	2,182	1998	Papua New Guinea
15,030	1792	Kyushu, Japan	2,144	1923	Tokaido, Japan
13,486	1771	Ryukyu Trench, Japan	2,000	1570	Chile
8,000	1976	Moro Gulf, Philippines	1,997	1946	Nankaido, Japan
5,233	1703	Tokaido-Kashima (J)	1,700	1766	Sanriku, Japan
5,000	1605	Nankaido, Japan	600	2005	Northern Sumatra
5,000	1611	Sanriku, Japan	119	1964	Alaska, USA

* Including earthquake's victims.
† The 40-meter waves from the tsunami destroyed 295 towns in the Sunda Strait in western Java and southern Sumatra.

few seconds of time and transmits the energy vertically to the surface. Table 4.5 gives the list of the most damaging tsunamis world wide in historical times. Tsunamis can occur not only in oceans but in all large water reservoirs: lakes, small seas such as the Mediterranean, the Black Sea or the Sea of Marmara, south of Istanbul. They are produced by sudden large underwater disruptions extending from bottom to surface, most often due to earthquakes. However, not all submarine earthquakes generate tsunamis: their magnitude should be above 6.5. They can also be due to submarine landslides and less often to undersea volcanic eruptions or pyroclastic flows, or meteorites hitting the ocean (Chapters 2 and 3) [38]. On 18 November 1929, a M7.2 earthquake 300 km south of Newfoundland triggered a submarine landslide displacing some 200 km^3. Other similar events have been reported such as the Storegga slide where, about 7,900 years ago, 1,000 km^3 of marine sediments collapsed from the continental shelf off the coast of Norway [39]. On 16 October 1979 one portion of Nice airport in France collapsed during extension work and generated a tsunami that killed several people in the nearby beach city of Antibes.

The force that maintains their waves throughout the water column is gravity. Their power is determined by the strength of the original disruption, the distance the waves travel, the configuration of the underwater terrain (if they occur deeply enough, the rock above absorbs their energy and much less is transmitted to the ocean surface), and the shape of the coastline. Their velocity across the ocean can reach several hundred kilometers per hour, and they travel several thousand kilometers in a few hours. For a 4-km deep ocean, the predicted wave speed is about 200 m/s and the waves can cross a 5,000-km ocean in about 7

hours. As they move very quickly, their energy is enormous. Their wavelength – the distance between two successive waves – can be several hundred kilometers. In the open ocean, tsunamis may be difficult to detect as the height of their waves might just be a few tens of centimeters near the point of disturbance. However, just like other kinds of waves, changes occur when the wave enters shallow waters near the coasts: their wavelength shortens to some 10 to 20 km, and as the fast water arriving from behind pushes the leading waves, their height increases, sometimes reaching several tens of meters.

The most sensitive areas of the planet are found along the Pacific coast where 85% of all earthquake-tsunamis occur because of the huge area of the Ocean and of the large seismic and volcanic activity of the Pacific tectonic plate. In comparison, the Atlantic Ocean counts only for 2% of all tsunamis recorded. It has been estimated that in the course of the last century, about 50,000 people died as victims of some 400 tsunamis. The most devastating event – and the first major one in the 21st century – followed the Sumatra–Andaman earthquake. It is also the most spectacular in history as it could be watched in near real time on TV screens all across the world. It killed some 230,000 people within just a few hours and left more than a million homeless [40]. This is equivalent to the effect of a 300-meter asteroid strike, similar to the strike that would be generated by Apophis (Chapter 3). Some scientists paradoxically called it an 'opportunity' for the future because it was the first tsunami for which tide-gauge measurements of high quality were available world wide and for which there were also, for the first time, multiple-satellite passes and altimetry measurements, allowing precise timing and wave height measurement in the open ocean. This unprecedented instrumental capability serves as an illustration of what ought to be done in the future to avoid a similar catastrophe.

4.5.2 The 26 December 2004 Sumatra tsunami

The Sumatra–Andaman M9 earthquake shocks started at 7.59 a.m. local time as reported by the Geophysics Institute in Djakarta, and raised a 1,200-km stretch of ocean floor by as much as 8 meters, displacing hundreds of cubic kilometers of water. The epicenter was localized at about 200 km west of Sumatra at a depth of 30 km. The Pacific Tsunami Warning Center (PTWC) in Hawaii issued the first information bulletin approximately 15 minutes after the earthquake. Although they correctly estimated the location of the epicenter, they underestimated the magnitude as 8.0, but nevertheless sent a warning of the possibility of a tsunami which, unfortunately, had no effect because of the deficiency in communication systems in the area. With a velocity of $\sim$500–800 km/h the waves took some time to reach the coasts. At 8.38 a.m. local time, 40 minutes after the first shocks, a first wave of 15 meters reached the beaches of Banda Aceh and of the Nicobar Islands, the closest to the epicenter, flooding 4.5 km inland and destroying all housing, even the most reinforced concrete structures that had bravely withstood the earthquake's shaking, and where most of the casualties were reported. Twenty minutes later it reached the islands of the Andaman Sea and the south of Sumatra. It took 11 hours to reach the southern coast of Africa. The

first ocean instrumental measurements from the real-time reporting tide gauge at the Cocos Islands – which was located approximately 1,700 km from the epicenter – were available about 3 hours after the earthquake [41], but were unfortunately of little use to those who had already died in Banda Aceh, Thailand and Sri Lanka.

Tide gauge and near real-time space-based measurements by the Jason-1 and Topex-Poseidon satellites, allowed us to observe the development of the first surface undulations from a few centimeters high near the epicenter, into the gigantic waves of destruction when they arrived at the shores. These data were used to build the numerical simulations and reconstruct the phenomenon and its global extension, as illustrated on Figure 4.14 [41]. They allowed the comparison to be made between measured wave heights and computed heights, evidencing the global character of the phenomenon. They revealed that wave amplitudes, directionality, and global propagation patterns were primarily determined by the orientation and intensity of the offshore seismic line source and, subsequently, by the trapping effect of the topography of the mid-ocean ridge. Seafloor topography is in fact the main factor determining the directionality of energy propagation. The coupling of global observations with numerical simulations helped to determine the principal factors affecting that portion of seismic energy that was transported thousands of kilometers throughout the world oceans by the tsunami waves. These simulations and their direct confrontation with local measurements proved, for the first time, that the paradigms used in the model could be applied in the future for prediction and public safety. However, forecasting the behavior of the waves as they approach and reach the shores is more difficult because of the multiple reflections caused by the local sea relief that strongly distorts their configuration and characteristics.

4.5.3 Forecasting tsunamis and mitigation approaches

Tsunamis are better understood than earthquakes but for those that result from an earthquake, their long-term forecasts are as difficult as for the quakes [18]. On the other hand, the principle of short-term forecasting is relatively simple, at least in theory: collect information on sea surface deformations in the middle of the ocean and transmit this information in real time to the specialists in dedicated centers for evaluation of the amplitude of the hazard and for decisions to evacuate the threatened areas. For trans-oceanic tsunamis, such as those in the Pacific or Indian Oceans, hours may elapse between an earthquake and the tsunami's arrival. The earthquake location and size can be quickly and accurately estimated by seismological observations, and the actual tsunami generation and propagation can be confirmed by an offshore sea level monitoring system such as bottom pressure gauges. As some tens of minutes elapse between the first shock and the arrival of the first killing wave, there should be time enough to escape the most dangerous areas, saving at least hundreds of thousands of human lives if not their housing. Numerical simulations as done with the MOST (Method of Splitting Tsunamis) model can also be utilized. For a tsunami of nearby origin,

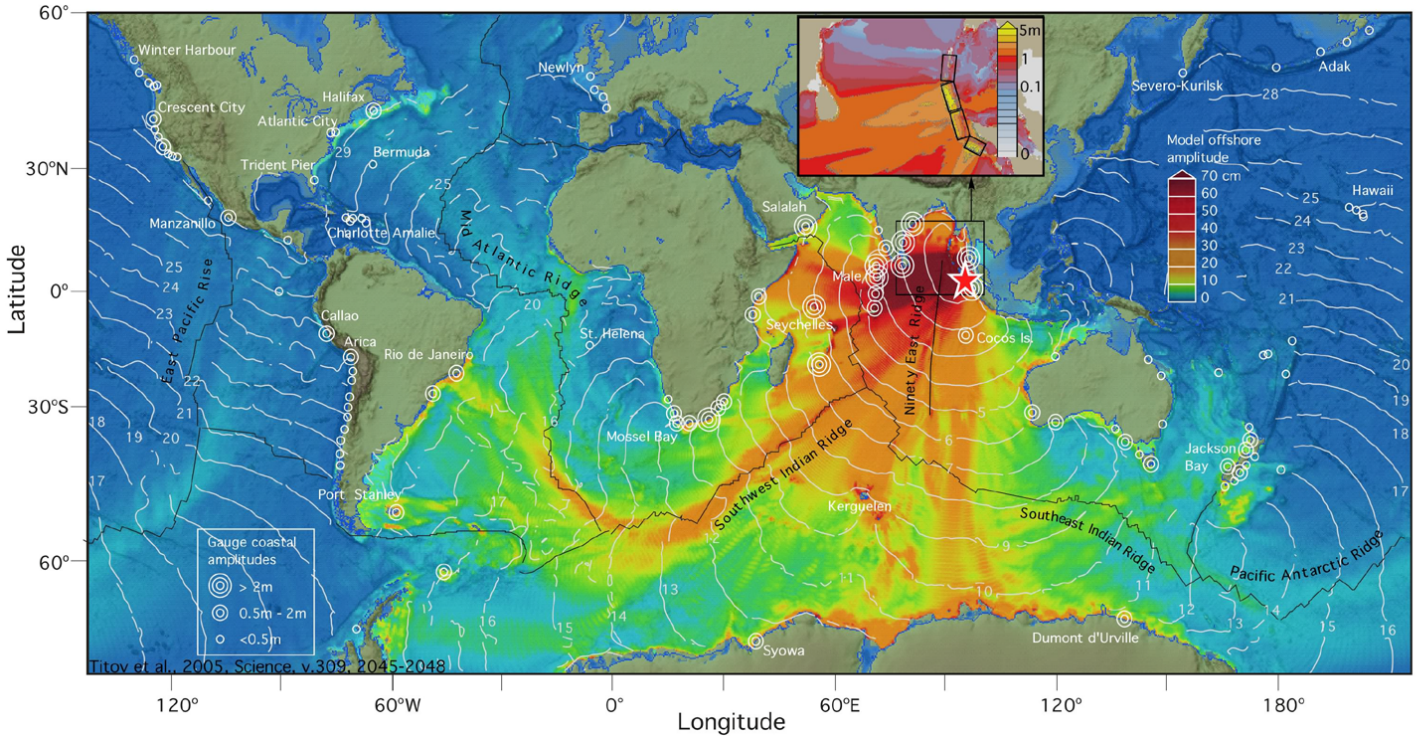


Figure 4.14 Global chart showing the energy propagation of the 2004 Sumatra tsunami calculated from the MOST (Method of Splitting Tsunamis) model of NOAA. Filled colors show maximum computed tsunami heights during 44 hours of wave propagation simulation. Contours show computed arrival time of tsunami waves measured in one hour slices. Circles denote the locations and amplitudes of the waves in three range categories for selected tide-gauge stations. The inset shows fault geometry of the model source and a close-up of the computed wave heights in the Bay of Bengal. Distribution of the slip among four sub-faults (from south to north: 21, 13, 17 and 2 meters) provides the best fit for satellite altimetry data and correlates well with seismic and geodetic data inversions. (*Credit*: Titov et al. [41].)

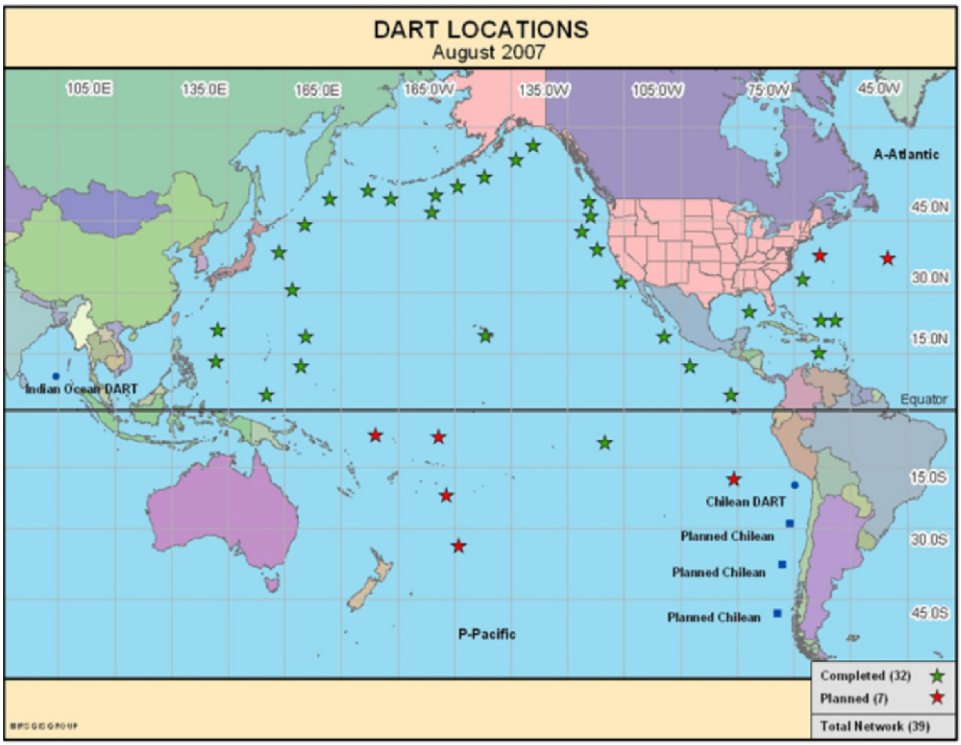


Figure 4.15 Pacific Ocean basin, showing locations of current (green stars) and planned (red stars) DART stations as of August 2007. (*Credit*: NOAA.)

the parent earthquake provides the most effective warning to coastal residents [42]. Monitoring the displacement of a fault potentially dangerous for tsunami generation with the help of the GPS will certainly become in the future an important element of any tsunami warning systems.

There exist a set of warning stations around the Pacific that measure any abnormal tide or wave activity appearing after a seismic event, such as the DART (Deep Ocean Assessment Reporting of Tsunamis) system of NOAA in the USA (Figure 4.15). DART covers the north Pacific, along the Alaska subduction zone, the most dangerous generator of tsunamis for the US west coast and Hawaii, as well as for Japan and other countries in the Pacific Rim. Because they are more often victims of these events the Japanese have developed one of the most efficient warning systems. These systems include seismometers, tsunameters – which detect the pressure of waves on the sea floor – tide-gauges near the shores and proper communication systems to send the warning signals to the scientists and to the officials in charge of evaluating the danger. In Japan, a tsunami warning based on seismology is issued 2 to 5 minutes after the earthquake, and is immediately relayed to coastal residents by the media or other methods (see Box 4.7). In that way, the Japanese have been able to drastically reduce the number of

their victims. Prior to the Sumatra tsunami, the Indian Ocean had no properly equipped warning system. Since then, the necessary infrastructure elements are gradually being put in place, in particular with the help of UNESCO's Intergovernmental Oceanographic Commission. However, to be effective, they need to be supported by prior assessments of tsunami hazards, land-use regulation and, last but not least, relentless education of coastal residents and tourists.

Similar systems are also foreseen in Europe – in particular, Portugal – and for Morocco. Of particular concern in this area is the most likely foreseen collapse of the unstable western flank of the Cumbre Vieja in the Canary Island of La Palma (impossible to forecast precisely unfortunately), which might create a 500-km^3 landslide and a tsunami that could span the whole Atlantic Ocean with waves of 10–20 meters along the east coast of North America, the western European coasts, and the Canary Islands themselves, where the waves might reach some 100 meters [38].

Box 4.7 Set of instructions following the first tsunami warning in Japan

1. Leave the seashore immediately and take shelter in a place of safety when a strong shake or a weak but long-duration slow shake has been felt.
2. Leave the seashore immediately and take shelter in a place of safety when a tsunami warning has been issued.
3. Acquire correct information from television, radio or via the internet.
4. Do not go to the seashore for bathing or fishing when a tsunami advisory or a tsunami warning is issued.
5. Do not relax until the warning is canceled as a tsunami may attack repeatedly.

Courtesy: Japan Meteorological Agency.

Unlike in Hawaii and Japan, many young people in South-East Asia, India and East Africa have never heard of a tsunami. A small amount of knowledge in the right place can indeed save many lives, as illustrated by the story of a 10-year-old British girl who had learned of tsunamis in school and warned her neighbors on the beach in Thailand on 26 December 2004 to move back from the shore and climb as high as possible. Overall, the Sumatra tsunami has eventually generated some optimism that its recurrence may not lead to a similar catastrophic balance in the future. Consequently, it is not unrealistic to say that in the next 100,000 years this human hazard will disappear, even though tsunamis will continue to exist.

4.6 Climatic hazards

Storms, floods and droughts are related to the Earth's weather and to the climate, and as such they are not 'purely' natural as the climate is more and more influenced by anthropogenic activities. Their energies are enormous both in terms of casualties and damage to property. As in the case of tsunamis, if we cannot avoid them, we might in future at least diminish, if not totally eliminate, their consequences as better weather and climate forecasting is achieved and mitigation measures are properly and timely implemented. For the time being, unfortunately, the cost of weather and climate-related disasters are rising. From 1980 to 2004, the global economic impact of such events totaled US$1.4 trillion (2004 economic conditions) [43].

4.6.1 Storms: cyclones, hurricanes, typhoons, etc.

Tropical cyclones form over the oceans where masses of moist rotating air continuously pick up more energy from the warm surface water until they become an organized system of thunderstorms. When the velocity of winds exceeds 119 km/h a cyclone is born. In the northwest Pacific, it is called a typhoon; in the Atlantic and northeast Pacific, a hurricane. The greatest attention is given to North Atlantic cyclones (hurricanes in that case) in spite of the fact that only 11% of the world's tropical cyclones occur there. This is because the USA is regularly affected by these disturbances, which are well documented, with the National Oceanic and Atmospheric Administration playing a key role. Aircraft fly regularly there to make precise measurements of the winds, and space observations are more systematic.

Tropical cyclones form only in special circumstances [44]. Grossly speaking, they work like a steam engine, converting thermal energy stored in the oceans into mechanical energy in the atmosphere. A cyclone removes heat and water from the ocean, cooling its surface as a result of evaporation, and transporting the moisture upward and poleward, thereby modifying the environment. The physical conditions for creating a cyclone are several. The temperature of the sea

Box 4.8 The names of cyclones

Since 1953, Atlantic tropical storms have been named from lists originated by the National Hurricane Center of the US National Oceanic and Atmospheric Administration. They are now maintained and updated by an international committee of the World Meteorological Organization. At first, all hurricanes had female names, but now, men's names alternate with women's names. Six lists are used in rotation. Thus, the 2006 list will be used again in 2012. In the event that more than 21 named tropical cyclones occur in the Atlantic basin in a season, additional storms will take names from the Greek alphabet: Alpha, Beta, Gamma, Delta, and so on. The other tropical zones (e.g. Pacific, Australia, etc.) have adopted a similar naming system.

from 50 meters below the surface upward must be above 26°C, because a minimum amount of ocean heat supply is needed as well as unstable air conditions and low-pressure atmosphere near the ocean surface. This explains why cyclones are found mostly in tropical zones, but some cyclones might also occur outside these zones. That was the case of the Catarina (see Box 4.8) cyclone (not to be confused with Katrina) which struck Brazil in March 2004, northeast of the Rio Grande do Sul State and south of the Santa Catarina State, at 28.5 degrees of latitude, the only such event ever recorded in the South Atlantic where none is expected to occur. Probably a more important condition for a cyclone to form is the vertical distribution of wind velocity, which must be regular. In scientific terms, the 'wind shear', which describes the different wind velocities at different altitudes, must be small to ensure that the convection cells that initiate the cyclone's vortex are not torn apart, and that the storm's heat is not dispersed. In fact, studies have shown that low wind shear is more important than high ocean surface temperature in modulating the formation of cyclones. The 'eye' of the cyclone, which corresponds to the central part of the vortex, is characterized by low and quiet winds and by the lowest pressures of the storm. It is delimited by a 'wall' where the most intense convective motions and the highest cumulonimbus clouds are found (Figure 4.16).

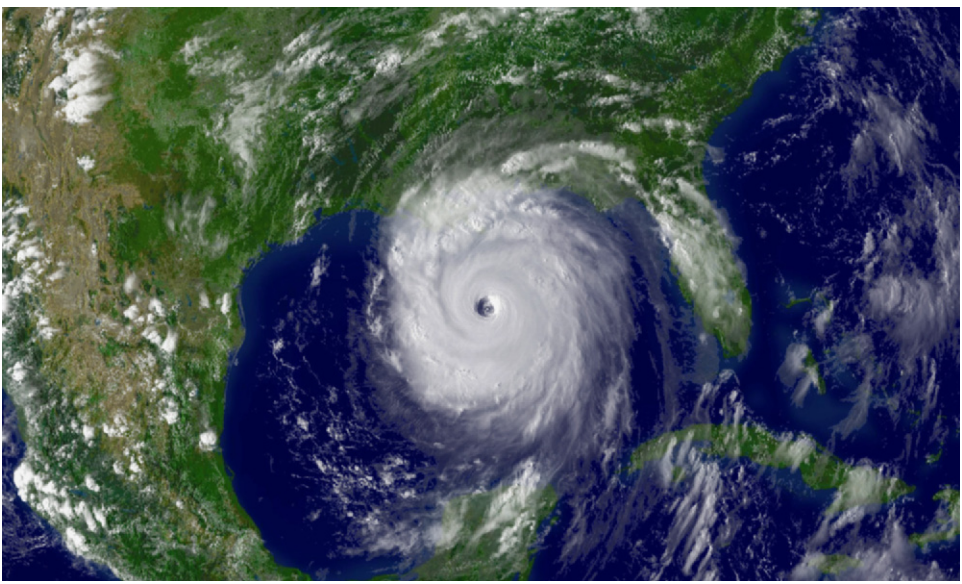


Figure 4.16 Katrina as observed with a US NOAA satellite before it hit New Orleans at the end of August 2005. Katrina was the third most intense storm ever recorded in the USA, with a central pressure of 918 millibars inside the very visible 40-km eye at the very center of the storm. It sucked so much heat from the Gulf of Mexico that, after its passage, the water temperature dropped drastically in some regions from 30°C to below 26°C. (*Source*: NOAA.)

Table 4.6 The Saffir–Simpson scale is used for characterizing the intensity of tropical cyclones. Between 119 and 63 km/h the storm is named a 'tropical storm', and below 62 km/h it is called a 'tropical depression'. For all categories, unanchored or temporary buildings such as mobile homes and poor constructions are severely damaged or definitely destroyed.

Category	Wind speed (km/h, knots)	Main damages
1	119–153, 64–82	No real damage to building structures.
2	154–177, 83–95	Some damage on buildings/considerable damage on shrubberies and trees.
3	178–209, 96–113	Severe damage on structures/evacuation possibly required.
4	210–240, 114–135	Very severe damage on buildings/evacuation required.
5	>240, 135	Some complete destruction/massive evacuation required.

The energies of cyclones are measured according to the wind velocity through the Saffir–Simpson scale described in Table 4.6. It is estimated that damage caused by cyclones rises in proportion to the cube of the wind speed [45]. Katrina (Figure 4.16) was category 5 over the Gulf of Mexico and had weakened to category 3 when it hit New Orleans on 29 August 2005.

The historical record of tropical cyclones is fairly poor but we know that they can be very bad: in November 1970, one of them killed more than 300,000 people in Bangladesh, as many as the Sumatra tsunami. In November 2007, cyclone Sidr hit the country again, resulting in more than 10,000 victims, and affecting some 7 million people. In 1780, 22,000 people died in La Martinique, St Eustasius and Barbados from a very violent hurricane. With wind velocities culminating above 260 km/h, the typhoon Saomaï was the most violent recorded since 1956, killing more than 400 people in China in August 2006. However, because these hazards are so destructive and because the USA is often hit by them, the death record is not the parameter most commonly used to evaluate their balance, but rather the costs induced for repair and reconstruction that have to be supported by insurance companies. Figure 4.17, which concerns the USA only, shows that the yearly costs of these disasters are about US$80 billion for an average number of four major events. Katrina and Rita have together killed 1,800 persons and the costs of their damages are above that average and close to US$85 billion or nearly half the total of all costs for 2005.

It has often been said that cyclonic activity is increasing in relation with global warming, and several attempts have recently been made to evaluate the evolution of that activity in the historical past. The instrumental record is obviously too recent, therefore the number of past cyclones is difficult to evaluate precisely. After the end of World War II, the North Atlantic archives are fairly reliable due to the use of aircraft, and since 1966 there exists a more routine set of observations from space. For the longer past, it is necessary to rely on 'paleotempestology', a discipline that uses geological signatures and deposits,

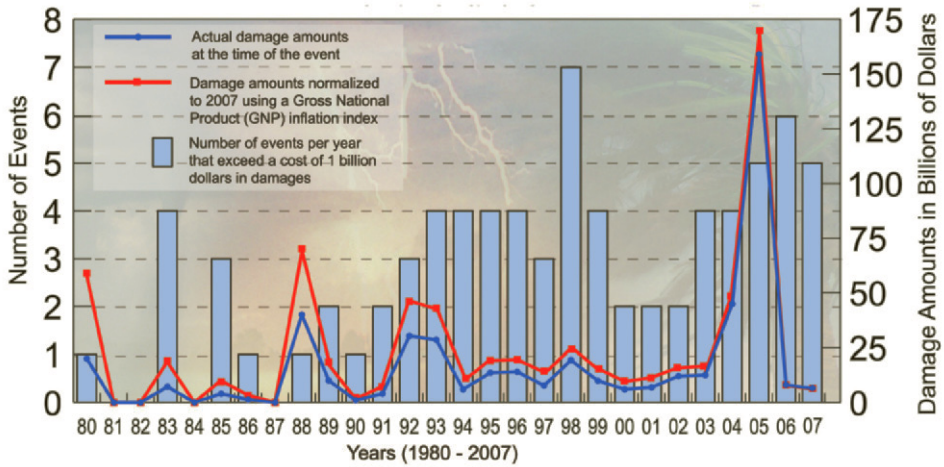


Figure 4.17 Annual costs of US weather disasters in billions of dollars as reported by NOAA for the period 1980–2006. (*Source*: NOAA/NESDIS/NCDC.)

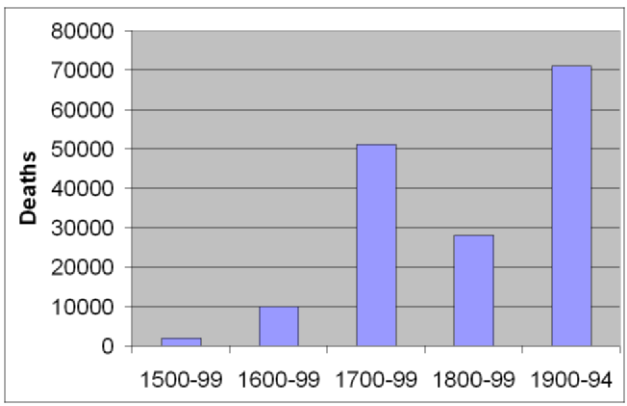


Figure 4.18 Deaths from tropical cyclones, shown in 100-year periods (except for 1900–1994) since 1500. (*Source*: NOAA–National Hurricane Center.)

such items as sediments and sand layers found at the bottom of coastal wetlands and lakes, which are deposited as sand dune barriers are washed over by the floods from the storm [46]. Bands in coral cores also help to reconstruct shearing winds and sea surface temperatures – two key parameters that influence the formation of cyclones [47].

Figure 4.18 shows the number of deaths from past cyclones since 1500. The increasing death toll with time most likely reflects the increase in population in cyclone-sensitive areas. However, the decrease observed in the 19th century is real and follows a period of high activity covering the past 300 to 150 years when sea surface temperature was in fact 2°C cooler than now. In the more distant

past, analyses of proxy data from Puerto Rico show a relatively frequent intense hurricane activity between 5,400 and 3,600 years before present, followed by a more quiescent period from 3,600 until roughly 2,500 years ago. From 2,500 to 1,000 years ago, another active period is evident from the past record which also corresponds to cooler-than-modern sea surface temperatures [46]. In oceanic basins other than the North Atlantic, figures are extremely difficult to evaluate: some storms that were evaluated as category 2 or 3 by the local authorities are now re-evaluated as category 4 or 5 after re-analyzing satellite data [45]. This makes it difficult to bluntly conclude that, globally, cyclone intensities are increasing until more records are available – a daunting task!

In the western North Atlantic Ocean, the frequency of intense hurricanes seems to be modulated by both the El Niño/Southern Oscillation, ENSO (Chapter 5) and by the West African monsoon. El Niño suppresses hurricane activity in the northern Atlantic by increasing the amount of wind shear and sinking air, something that hurricanes do not like. In effect, hurricane activity appears to be lower during the El Niño events [46]. A relation is also noticed between the amount of precipitation in tropical Africa – directly connected to the strength of the West African monsoon – and the strength of hurricanes: the stronger the monsoon, the larger the hurricane activity in the North Atlantic. A better understanding of how these climate patterns will vary in the future is therefore required to predict hurricane activity more accurately.

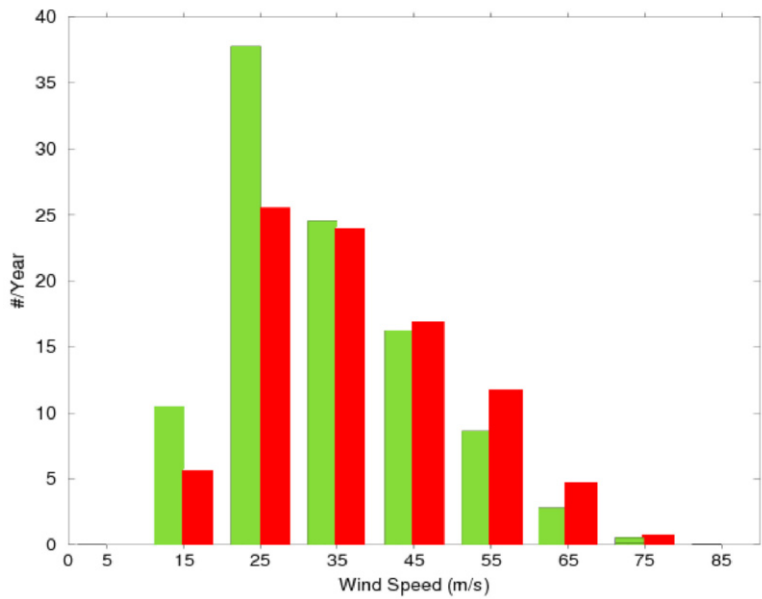


Figure 4.19 Predicted changes in the maximum attained intensity distributions for northern hemisphere tropical cyclones between the end of the 20th (green vertical bars) and 21st century (red bars): the number of cyclones decreases with an increase in global warming, but the intensity of cyclones increases. (*Credit*: L. Bengtsson [44].)

Remarkably, the accuracy of predictions is continuously increasing and this is one of the great achievements of international meteorology [44]. The use of super-computers in Europe and Japan in the assessment of global warming and of its consequences over the next century, is delivering some surprising conclusions. An increase in the sea surface temperature larger than 2°C at the end of the 21st century will not result in a larger number of tropical cyclones, but rather in a reduction by 12%. This is most probably the result of a general weakening of the large-scale atmospheric circulation, as the higher sea surface temperature means more water present in the atmosphere, hampering the formation of vortices. Nevertheless, there would be an annual increase of tropical cyclones of category 3 and above, which is what might logically be expected (Figure 4.19). This is true for all regions of the globe where these cyclones exist, with the drop in the number of cyclones being particularly striking in the Indian and the West Pacific Oceans. These conclusions do not contradict the continuous trend observed over the past 30 years for stronger cyclones occurring in every ocean, and closely tracks the rise in sea surface temperature. Even though they represent the outcome of models and can therefore be criticized on the basis of the parameters and of the assumptions that enter into these models, this is in no way an excuse to ignore the reality of global warming [48].

However, it is one thing to have good forecasts and another to use them efficiently and prepare for the worst. Even though the USA had one of the most advanced centers (the National Hurricane Center, located in Florida) that could forecast storm paths with an accuracy that has improved by 50% in the past 30 years, and despite the fact that US meteorologists had warned that a catastrophic event would threaten New Orleans, such a catastrophe actually happened on 29 August 2005 with Katrina. Certainly, the scientists or the forecasters have a responsibility to inform the public and the authorities of the imminence of a disaster, but it is then up to these authorities to react. In the case of Katrina, it has been known for a long time that the water circulation system that kept New Orleans dry was designed to withstand hurricanes up to category 3, but not above. The necessary investment to adapt this system to the long foreseen catastrophe was never made, probably because the scientific predictions based on models were neither considered nor accepted as being serious enough by the politicians [49]. Katrina offers a clear demonstration that science alone cannot guarantee protection if the responsible people or institutions do not apply the proper mitigation policies. Can we imagine the situation in countries neither as rich nor as well educated and scientifically developed as the USA? Let's hope that it will not take 1,000 centuries to find a solution to a problem that should not exist, provided science predictions and mitigation measures are implemented with a firm will.

4.6.2 Floods

Cyclones and storms are major climate and weather-related causes of floods. The catastrophic flooding of New Orleans represents one striking example of the amplitude of such disasters. But floods also have other causes.

Historical megafloods

For many years in the past, major floods or megafloods have been associated with large ruptures of the Earth's surface, with earthquakes and their subsequent tsunamis. They have reshaped continents and the distribution of water between oceans and seas. That was the case when the evaporative conditions in the Mediterranean –then a desert – ended some 5.33 million years ago at the boundary between the Miocene and the Pliocene epochs, following a global rise of sea level, when water from the Atlantic Ocean re-entered the Mediterranean from the Gibraltar Strait, ending the Messinian salinity crisis (see below). At the same time, a major flood occurred in the Bahamas Platform, supporting the assumption that this huge sea-level rise was global [50].

Closer to us in time, megafloods occurred twice in England between 450,000 and 180,000 years ago, shaping the Channel between England and France and making England an island. This is the conclusion of a 25-year ocean-floor mapping by a group of British geophysicists which evidenced the existence at the bottom of the Channel of a large valley with a bedrock floor marked with streamlined islands scattered down along its axis – features that are characteristic of very large amount of water pouring down [51]. These megafloods are assumed to result from the breaching of a rock-dam located near the present position of the Dover Strait that discharged into a huge lake located where the North Sea is now. The flow lasted for months with discharges reaching several million cubic meters per second and carving the 50-meter deep valley into the bed rock.

Using seismic waves for imaging the layers of sediments at the bottom of the Black Sea, a group of American and Russian oceanographers are claiming evidence that, about 7,500 years ago, a megaflood suddenly filled the Black Sea, which was then a fresh water lake, to its present level – an assumption that is strongly supported by the presence of a thin, uniform dusting of sediment, consistent with a geologically instantaneous refilling of the Sea. In addition, radiocarbon dating of the shells of the first salt-tolerant mollusk invaders from the Mediterranean yielded an age of 7,550 years before present, plus or minus 100 years. Finally, seismic probing has shown that the hard-rock basement beneath the sediments filling the Bosporus channel lies at a depth of nearly 100 meters rather than 35 meters as had previously been thought, which also supports the scenario whereby the floods have cut a very deep channel through the sediments down to the bedrock, allowing the water to enter even more quickly. All these facts seem to confirm the occurrence of a deluge about 7,500 years ago which some [52] have rapidly – too rapidly maybe – associated with the Deluge described in the epic poem of Gilgamesh from Babylonia, one of the earliest literary works, and two millennia later in the Genesis chapters of the Old Testament (Figure 4.20). This appealing association is not supported, however, by a more serious analysis, notwithstanding the question of how the memory of that event could have been retained orally for several millennia as writing did not exist 7,500 years ago.

Seasonal climate and weather-related floods have marked the history of humanity over many centuries in the past. In many temperate regions around



Figure 4.20 Mosaic in the San Marco Basilica in Venice illustrating the Deluge.

the world, spring floods are common. They are usually associated with snow melt or thunderstorms. However, floods can occur at any time of the year depending on location, their timing being largely dependent on climate and seasonal weather patterns. The annual cycle of flooding and farming was a significant phenomenon not only for the Egyptians of the Nile river and the Mesopotamians of the Tigris and Euphrates rivers, but also the Indus, the Ganges and the Yellow river. There, the flooding of agricultural land also brought some benefits, making soil more fertile and providing nutrients wherever deficient. As such, these floods cannot be seen as major hazards. However, the continual extreme weather and climate conditions are making their impacts more difficult to cope with, as is witnessed more and more frequently.

Reshaping the planet, killing people and displacing entire communities and large sets of population clearly make floods very unpleasant and dangerous hazards. This being said, however, they have not hampered the existence of civilization and none has been identified as being the cause of mass life extinction.

Present-day floods

Present-day floods are the most frequent type of disasters world wide. Although they are usually less common in dry environments and highlands, as rain falls and snow melts nearly everywhere on Earth, few places are spared by them. According to the WHO, more than 2.8 billion people have been affected by floods in the course of the 20th century: they were either drowned, or died from



Figure 4.21 The 'acqua alta' in Venice. This weather-related phenomenon is made worse by the hydrological work done to enlarge some of the channels in the city.

the diseases caused by the floods, or were injured, or lost their homes. Seven of the 10 deadliest flood disasters in that period have occurred in China, where more than 6 million people died from drowning, starvation and disease during the three biggest floods in 1931, 1939 and 1959. More recently, in February and March 2000, Mozambique was victim of catastrophic flooding caused by heavy rainfall that lasted for five weeks and caused rivers to break their banks. The situation was intensified when Cyclone Eline hit the same area on 22 February 2000. Two million people were affected by the floods, 25,000 were made homeless and about 800 were killed, 1,400 km of arable land was affected and 20,000 head of cattle were lost. It was the worst flood in Mozambique in 50 years. In August 2007, India, Pakistan, Bangladesh and Nepal witnessed the worst monsoon in recent history, with more than 2,200 deaths and 30 million people affected.

Sea floods are major events leading to major disasters as many communities are located near the coasts. They represent a major threat around the world in particular to populations living in equatorial countries such as Bangladesh, but not only there. They also affect Europe, and Holland in particular. The most vulnerable areas are the flat plains and the sea shores in very populated areas where the largest number of casualties are found. This is constantly increasing because the population is increasing, the environment is degrading, and planning, land-management and preparedness are poor or non-existent. Floods

also occur when sea water is being blown inland by the winds such as when the high waters regularly flood the city of Venice in Italy – the 'acqua alta' which results from the combination of rain, high tide and the sirocco blowing towards the city (Figure 4.21). A similar phenomenon could affect other cities, such as New York.

Forecasting floods

Proper forecast and mitigation measures should result in making floods in the future more of an embarrassment than a real hazard. The first and most obvious measure is to understand their causes. One of the main causes is related to the climate and, for some parts of the world, to El Niño. Obviously, we must avoid destabilizing the climate any furher, and slow down global warming, as discussed in the next two chapters. Another cause is deforestation as flooding is less likely to occur on forested grounds than on bare soils: leaves receive water that can later evaporate, while roots create openings in the ground into which water can seep. In deforested areas, on the contrary, rain does not penetrate the soil, but washes it away and may cause sudden floods. Of course the solution is to avoid cutting trees or, if the damage is already done, to replant them.

The ability to forecast flooding is limited to the time during which changes in the hydrological conditions necessary for flooding to occur have begun to develop. In some areas, such as India, floods can be relatively predictable events. In most areas, however – especially those affected by mid-latitude cyclones – floods can be difficult to predict more than 24 hours in advance. The formulation of a forecast for flood conditions requires information on current hydrological conditions such as precipitation, river stage, water equivalent of snow pack, temperature, soil conditions over the entire drainage basin, as well as weather reports and forecasts. In small headwater regions the relatively rapid rate of rise and fall makes the period of time above the flood stage relatively short. In lower reaches of large river systems, where rates of rise and fall are slower, it is important to forecast the time when the various critical stages of flow will be reached. Reliability of forecasts for large downstream river systems is generally higher than for headwater systems. Higher reliability weather forecasts and repeated space observations should provide the best warning information for pre-disaster management. Remote sensing of the flooded areas and their evolution prove most valuable for post-flood analysis and development of the proper mitigation approaches.

Presently, the amount of information that is required, the data collection network that is necessary for collecting the information, the technical expertise necessary for interpretation, and the communication system needed to present timely information to potential victims are services that many poor and developing nations find difficult to provide. The WMO, through its World Weather Watch and Global Data Processing System, hopes to coordinate efforts to improve forecasting world wide. This is especially important (but difficult) when conditions creating floods lie outside the national boundaries of the downstream region, as in the case of Mozambique (see below).

Mitigation

How can we protect ourselves from floods? The answer is to a large extent in our hands. A good example is close to us in Holland, many times the victim of dramatic floods in past centuries. After the North Sea flood of 1953 of the southwestern part of the Netherlands, the Dutch authorities did not wait long to make the right investment, building the largest and most elaborate flood defenses and systems of dykes ever, thereby saving many lives by protecting their low lands from catastrophic floods. The Dutch had already built one of the world's largest dams in the north of the country: the Afsluitdijk whose closure occurred in 1932.

Dams offer a solution to the regulation of river flows. When a dam is built, the flow of the river is changed, usually keeping a somewhat constant flow rate all over the year (contrarily to what happens in naturally flowing rivers). This leaves the river catchments basins open for an annual occupation by humans and land animals and plants. However, dams require careful construction and management. Frequently, they eventually reach their limit and have to release more water than they would otherwise release. For example, the catastrophic Mozambique floods of 2000 were rendered more severe because of improper management. The opening of the dams in Botswana, Zimbabwe and South Africa increased even more the Limpopo and Zambeze river torrents, putting the Cahora Bassa Dam in Mozambique under strong stress. When it was decided to open the Dam in the fear that it would not withstand the supplement amount of water, those who survived the previous floods had then to face the new floods of waters pouring downstream from the Dam. The Three Gorges Dam under construction in China will hopefully regulate the Yangtse river but there are also concerns about the future of the several million people who will be displaced by the rising waters, the loss of many valuable archaeological and cultural sites, as well as the effects on the environment (Chapter 8). Nevertheless, it is remarkable that flood casualties in China have dropped throughout the 20th century, in part because of investment in protection systems and in part because of appropriate evacuation plans: in the floods of the 1930s and 1940s , 4.4 million people died; in the 1950s and 1960s, that number had dropped to 2 million and by the 1970s and 1980s it was less than 15,000.

In tropical countries, one strategy is to restrict the deforestation or to reforest lands that have been cleared and lie upstream of a potential flood region [53]. Some feel, however, that forests offer no defense against extreme floods, and that money should be better invested on other measures such as discouraging human settlements in flood plains. Nevertheless, such a strategy has several other positive effects on the limitation of wild fires, the conservation of biodiversity and the slowing of global climate change by increasing the capacity of CO_2 absorption and the development of sunlight-reflecting clouds through large-scale evapotranspiration.

4.6.3 Droughts

Droughts are characterized by extended periods when the amount of water falling on or arriving in an area of the Earth is lower than it normally is,

putting plants, animals and human beings in a critical situation of stress and survivability. The quantity of water on Earth being constant – until we reach global temperatures around 70°C which would evaporate and dissipate water into the upper atmosphere from where it would then dissociate and escape into space, transforming our planet into a second Venus (Chapter 9) – droughts should in principle be compensated by higher than normal precipitations, but not necessarily in the same area. Droughts can in effect happen everywhere on the planet. They are a normal, recurrent feature of the climate with some areas being more sensitive than others. But when they affect the same area several years in sequence, they result in grave ecological catastrophes and in severe problems of adaptation, of food shortages, of health and survivability, and possibly of population migrations. Their impact in the 20th century is also measured in large numbers of affected people: more than 2.2 billion according to the WHO, of which more than 10 million died of famine or disease.

Historical droughts

Droughts are mostly a climatic phenomenon but can also result from continental modifications and redistribution of the water reservoirs. One such example is the drying out of the Mediterranean, or Messinian Salinity Crisis already mentioned. Between 6 and 5 million years ago, the Mediterranean became isolated from the Atlantic Ocean probably as the consequence of a tectonic closure of the Straits of Gibraltar and of a climate-related fall in sea level. It dried out and was replaced by salty lakes recharged by deep canyons and rivers, killing the normal marine fauna that had existed previously. That crisis, we have just noted, ended by an abrupt flood when the Atlantic Ocean waters poured down through the Straits of Gibraltar like the Niagara Falls into that kind of 'Mediterranean desert', 5.33 million years ago.

In more recent times, during the second half of the twentieth century, the harshest drought affected the Sahel – that region which extends from eastern to western Africa in the sub-Saharan latitudes and includes such countries as Mauritania, Senegal, Mali, Burkina Faso, Niger, Chad, Nigeria and Sudan, representing about 20% of the landmass of Africa and where we find the poorest regions of the world. More than 10 million agricultural people, farmers and shepherds had to migrate and drastically change their ways of living, going from country to cities, and multiplying the number of inhabitants there by more than an order of magnitude [54]. Amazingly, by the end of the 1990s, rain was again falling in the Sahel, but in too great a quantity, as the region has witnessed drastic historical floods giving an unpleasant demonstration of the adage that 'what goes up must come down'!

But the worse is probably to come in Asia as the population is increasing and, according to the United Nations, the Himalayan glaciers, which are the source of the biggest rivers in India and China – the Ganges, Indus, Brahmaputra, Yangtse, Mekong, Salween and Yellow river – could vanish by 2035 as the result of global warming. If this occurs, the populations of India, China, Pakistan, Bangladesh

and Nepal, representing more than 2.4 billion people, will find themselves under the pressure of huge floods that will then be followed by harsh droughts.

Causes of droughts

Weather-related droughts are the result of reduced amounts of atmospheric water vapor and rainfall due to several different factors, such as abnormally long-lasting anticyclonic conditions, prevalence of continental over oceanic winds, El Niño, deforestation, and agricultural practices – in particular, intensive farming. In 2005, parts of the Amazon Basin experienced the worst droughts in 100 years which, adding their effect to deforestation, are placing the basin under high risk. The Sahel problem, originally the result of low levels of rainfall and high temperatures, was worsened by the deterioration of the soil, which was no longer protected by vegetation that was exhausted by small-scale agriculture and cattle feeding, leading to desertification. There are some hopes that the complex mechanisms of the African monsoon will be better understood. Unfortunately, some droughts are essentially the effect of poor management of land and water resources, as in the case of the Aral Sea (see Chapter 8).

Effects and consequences of droughts

Long-lasting droughts can have very serious environmental economic as well as social and political consequences. The effects on agriculture, the death of livestock and soil erosion are particularly harsh. Among them, desertification gives rise to much concern. It results from the soil slowly losing its ability to grow plants, followed by a period of rapid deterioration which ends up by the soil's inability to retain nutrients and water, thereby leading to plant death. This is a long process which is difficult if ever possible to reverse. Approximately 65,000 km^2 of the Earth's surface is turned into desert each year. Since droughts are also triggered by high temperatures, they are often associated with fires which are being witnessed more and more frequently in all parts of the world. Fires destroy the forests and open even further the way to the process of desertification.

The effects on humans can be substantial. Heat waves, for example, are frequently associated with droughts. During droughts, the cloud cover is nearly non-existent and sunlight heats the soil's surface, evaporating what remains of its moisture. The heat from the ground is directly transferred into the air which becomes even hotter. Heat waves are responsible for more deaths than any other type of weather disaster. The most frequent causes of death are cardiac arrest, strokes, dehydration and respiratory diseases. The heat wave of 2003 in France, when temperatures reached values of more than 40°C for seven consecutive days, caused the deaths of some 20,000 elderly people, with an equivalent number in Italy (Figure 4.22). In poorer countries such as Africa, Asia, and even Australia – about 90% of Australia is arid or semi-arid – famine is the most drastic consequence of droughts, forcing populations to move to other areas. The combination of drought, desertification and overpopulation is one of the causes of the Darfur wars as the Arab Baggara nomads, who are searching for water for

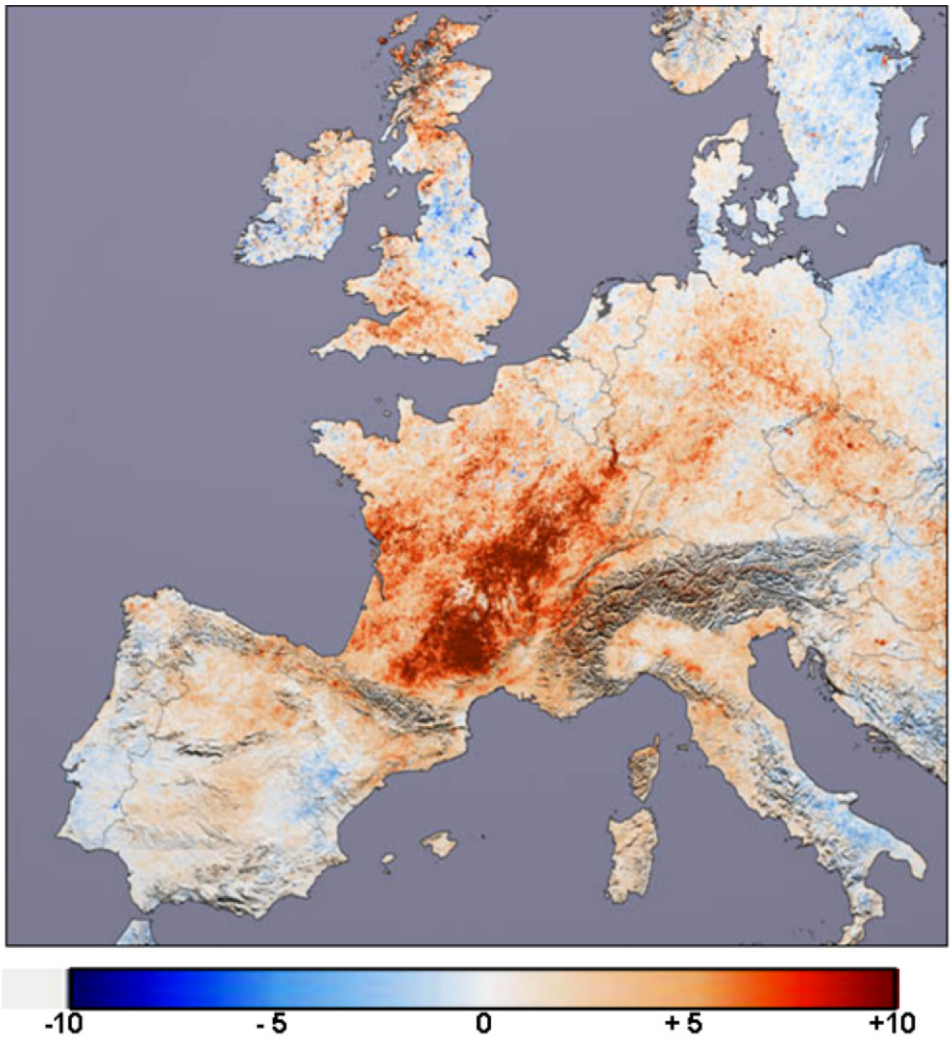


Figure 4.22 2003 heat wave temperature variations in comparison to normal temperatures in Europe. The color scale indicates the temperature anomalies in degrees Celsius. (*Source*: NASA.)

their herds, are forced to move further south into less affected lands mainly occupied by farming populations, that are not Arab.

Furthermore, even though droughts produce less sensational pictures of destruction than some other hazards we have discussed, their economic impact can be particularly disastrous in poor countries, but not only there. In the United States, for example, droughts cost between 6 and 8 billion dollars each year [54]. When water becomes scarce in rivers and lakes, not only do the various forms of aquatic life disappear but ships and barges may no longer operate, preventing the

transport of goods and raising their costs, as other and more expensive means of transportation have to be used. Hydraulic electricity production may also be affected.

Warning and mitigation against droughts

Further possible progress in climate forecasting represent the best hope today of forecasting these nasty hazards. Earth observation satellites are now capable of measuring the water levels through different techniques, in particular the modification of the Earth's gravity under the weight of the water mass, the measurement of soil moisture content and evaporation rates (Chapter 10). Satellite images are also essential in monitoring deforestation, receding glaciers and crop growth. A combination of this information with data on precipitation and temperature and other factors characterizing the climate, offers some prospects of predicting the length and intensity of a drought in the future.

The obvious straightforward mitigation measure is to bring the water back into the regions where it is lacking. One inventive solution that has been proposed in the last 50 years of the 20th century was to provoke rainfall by artificially seeding clouds. The efficiency of that technique is still debated, however: some studies suggest that cloud seeding has no effect as it is difficult to know whether a cloud would have produced rain without the seeding; others credit the technique with an increase in precipitation by 5 to 20% only [54]. Another solution, probably within reach in the very near future, would be to proceed with sea-water desalinization, a process discussed in Chapter 8. A parallel approach consists in the continuous monitoring of rainfall and current water usage levels with ground-based measurements and space observations, ensuring that over-usage is not putting the reserves at risk. The implementation of proper water management procedures can make the effects of droughts less severe. These are water conservation, the collecting and storing of rain water, recycling and irrigation, although this is not necessarily the most efficient way of keeping the water where it should stay and of avoiding the depletion of aquifers. Restrictions are a last solution only if unavoidable. Other measures of more local effect consist in acting on the farming itself, using drought-resistant crops and planting rows of trees or shrubs to lower the effect of winds and retain the water in the soil.

4.7 Conclusion

Some of the most deadly natural hazards discussed here – volcanic eruptions, earthquakes and tsunamis, hurricanes, floods and droughts – are beyond the human capacity to manage. Among these, volcanic eruptions can be credited in the history of Earth to be the likely cause of some dramatic death records and life extinctions. All the others, even though they can result in a large number of direct or indirect causes of death, are far from representing potential global threats that might result in the disappearance of civilization. Furthermore, with

some clever and politically supported mitigation approaches, their death records can be substantially lowered if not totally eliminated.

In the past and at present, it is striking that the poor countries are the most vulnerable to all these disasters. Population growth, which is a fact of most developing countries, forces more and more people to live in disaster-prone housing which make earthquakes, cyclones and floods as well as wildfires and disease propagation more deadly. The Sumatra–Andaman tsunami and the disaster of Katrina in New Orleans strikingly put into evidence the crucial importance of the handling of disaster management which is today so badly failing.

Nature does not know which countries are poor and which countries are rich. Human beings do, however. They have the brains and possess the modern tools that have the power and the capability of lowering the impact and the effects of the ups and downs of Nature. This is clearly the case of space systems which are becoming so important in the management of all hazards discussed in this chapter and which we describe more extensively in Chapter 10. The level of development of vulnerable populations is a factor that modulates the consequences of natural hazards. This is even more striking for diseases and deaths resulting from environmental problems (polluted water, poor sanitation, etc.) where studies conducted by the WHO show that they could be reduced by nearly a factor of 3, saving some 13 million lives per year, much more than the number of deaths due to AIDS, for example. Many of these health problems do not necessarily require a medical solution.

Mitigation measures are the most critical for lowering the level of danger of natural disasters and for preparing populations to learn how to deal with them, before, during and after they occur. Prevention implies both scientific rigorous analysis of past phenomena and the understanding of the most influential parameters characteristic of each category of disaster, as well as the development of social and political measures such as research, education, the strengthening of infrastructures, and better cities, land and soil management approaches – in particular, in regions that are prone to flood and drought-related hazards. One of the main difficulties is the 'myopia' of some governments to timely invest on protections against events of low probability. There again, scientific preparation and education by the scientists, not only of the people but also of the politicians, is a key element in the whole process. This, indeed, places a heavy responsibility on the scientists. They should more systematically abandon the 'purity' of their legendary ivory tower, the attitude of strict objective fact-finding and the predominant usage of highly specialized articles, and start using a more common and appropriate language that the public, the media and the politicians can properly understand. This, however, should not be at the expense of an indispensable rigorous scientific analysis. In that way, scientific results and their interpretations in the broadest context might have a chance of resulting in actions being taken.

At the global level, risk and disaster management require multilevel world governance systems that can enhance the capacity of coping with uncertainty

and surprise by mobilizing the appropriate sources of mitigation, but also post-disaster and resilience. There is some hope that right steps in that direction are already being taken. The United Nations Development Program (UNDP) has a Disaster Reduction Unit which is built on sharing information and knowledge with public authorities around the world, and through them raising public awareness. Some have suggested setting up an International Panel for Natural Hazard Assessment [55] which would function like the Intergovernmental Panel on Climate Change (Chapter 11) and extract, from the enormous amount of science data and models, some generally accessible messages and summaries of their most up-to-date findings for governments, local planners and populations to understand and, ultimately, on which to act. Applying such an approach would most likely have saved a substantial percentage of the 300,000 deaths in Asia in December 2004.

We should not believe, however, that in that connection the poor countries are in a worse situation than the rich in politically reacting to properly-phrased scientific advice. Some of the richest countries simply or hypocritically close their eyes and cover their ears when their most immediate short-term interests are at stake. This is clearly the case for the now utterly important, most concerning, and globally crucial issue of global warming. The changing climate of the Earth is certainly one of the most obvious demonstrations of the complexity of the whole Earth system, and we dedicate the following two chapters to this fundamental component of the future evolution of our planet.

4.8. Notes and references

[1] Mathers, C.D. and Loncar, D., 2006, 'Projections of global mortality and burden of disease from 2002 to 2030', *PLoS Medicine* [online journal], 3 (11), 442 (http://medicine.plosjournals.org/perlserv/?request=get).

[2] Willmoth, J., 1998, 'The future of human longevity: a demographer perspective', *Science* **280**, 395–397.

[3] The oldest human on record, Jeanne Calment from Arles, France, died in 1997 at the age of 122, probably from heart disease.

[4] Kirkwood, T.B.L., 2005, 'Understanding the Odd science of aging', *Cell* **120**, 437-447.

[5] Balaban, S. et al., 2005, 'Mitochondria, oxidants and aging', *Cell* **120**, 483–495.

[6] Proust, J., 1999, *Tout savoir sur la prévention du vieillissement*, Favre Ed., Lausanne, p. 217.

[7] Proust, J., 2007, Private communication.

[8] Saint-Pierre, J., 2002, 'Problèmes de limites en statistique', Report of the Centre Interuniversitaire de Calcul de Toulouse, 118, Route de Narbonne 31062 Toulouse Cedex 04 (Joseph.Saint-Pierre@cict.fr), p. 26.

[9] Simkin, T. et al., 2001, 'Volcano fatalities – lessons from the historical record', *Science* **291**, 255.

[10] Siebert L. and Simkin, T., 2002, *Volcanoes of the World: An Illustrated Catalogue of Holocene Volcanoes and their Eruptions*. Smithsonian Institution, Global Volcanism Program Digital Information Series, GVP-3.
[11] Rampino, M.R., 2002, 'Super-eruptions as a threat to civilizations on Earth-like planets', *Icarus* **156**, 562–569.
[12] Bourdon, B. et al., 2006, 'Insights in the dynamics of mantle plumes from uranium-series geochemistry', *Nature* **444**, 713–717.
[13] Wilson, M., 2006, 'Tectonic plate flexure may explain newly found volcanoes', *Physics Today* **59**, 21–23.
[14] Church, J.A. et al., 2005, 'Significant decadal-scale impact of volcanic eruptions on sea level and ocean heat content', *Nature* **438**, 74–77.
[15] Cazenave, A., 2005, 'Sea level and volcanoes', *Nature* **438**, 35–36.
[16] Hill, D.P. et al., 2002, 'Earthquake–volcano interaction', *Physics Today* **55,** 41–47.
[17] Benioff, H. et al., 1961, 'Excitation of the free oscillations of the Earth by earthquakes', *Journal of Geophysical Research*, **66**, 605–619.
[18] Stevenson, D., 2005, 'Tsunamis and earthquakes: what physics is interesting?', *Physics Today* **58**, 10–11.
[19] An inverse analysis of the time arrival of seismic waves together with the frequencies of the waves yield (indirect) information on layers which are physically inaccessible. Interestingly, the same techniques and equations have been used by solar astronomers to infer the internal properties of the Sun, measuring the vibrations of its surface, as induced by the convective and turbulent motions of the upper layers of our star. Through what is now called helio-seismology, the temperature, chemical composition and motions of the solar interior have been determined down to only a few per cent of the solar radius. This is an area of science where astronomers and geophysicists tend to cooperate.
[20] Pidwirny, M., at the University of British Columbia Okanagan. Copyright 1999–2007 Michael Pidwirny.
[21] McCaffrey, R., 2007, 'The next great earthquake', *Science* **315**, 1675–1676.
[22] It is said that the 1906 San Francisco event has triggered modern earthquake research. Andrew Lawson from the University of California, Berkeley, mapped the San Andreas Fault and established the 'elastic rebound model' explained on page 121 (see Marshall, J., 2006, '100 years on, you'd think San Francisco would be ready', *New Scientist*, 15 April 2006, 8–11).
[23] A special issue of *European Review* Vol. **14**, May 2006, Edited by Cambridge Univ. Press, has been dedicated to the Lisbon earthquake.
[24] Brunious, C. and Warnera, A., 'Earthquakes and Society', http://www.umich.edu/~gs265/society/earthquakes.htm
[25] Yeats, R.S. et al., 1997, *The Geology of Earthquakes*, Oxford University Press, VI, p. 568.
[26] Tarbuck, E.J. and Lutgens, F.K., 1996, *Earth: An Introduction to Physical Geology*, Prentice Hall: New Jersey, XVII (5th Edition), p. 605.

[27] Kanamori, H. and Brodsky, E.E., 2001, 'The physics of earthquakes', *Physics Today* **54**, 34–40.
[28] An M9 earthquake accounts for about 20 meters of slip on the boundary between two plates which converge at 0.02 to 0.1 meter per year; thus, an average time between earthquakes is about 200 to 1,000 years, assuming all the slip is by M9 events. If the slip occurs through smaller quakes, this interval might most likely be longer [19]. According to an updated estimation of their age and rate dependence, major quakes of M ~ 9 all occurred at sub-duction zones where the sub-ducting plate is less than 80 million years old and where the plate convergence is between 30 and 70 mm per year.
[29] Feldl, N. and Bilham, R., 2006, 'Great Himalayan earthquake and the Tibetian plateau', *Nature* **444**, 165–170.
[30] Subarya, C. et al., 2006, 'Plate-boundary deformation associated with the great Sumatra–Andaman earthquake', *Nature* **440**, 46–51.
[31] The Global Seismographic Network (GSN), composed of 137 ground-based stations is designed to detect motions within microns, distributed world wide and run by the Incorporated Research Institutions for Seismology (IRIS) in collaboration with the USGS, is one of the most useful tools for earthquake detection.
[32] Olsen, E.L. and Allen, R.M., 2005, 'The deterministic nature of earthquake rupture', *Nature* **438**, 212–215.
[33] Han, S.C. et al., 2006, 'Crustal dilatation observed by GRACE after the Sumatra–Andaman earthquake', *Science* **313**, 658–662.
[34] Ouzounov, D. and Freund, F., 2004, 'Mid-infrared emission prior to strong earthquakes analysed by remote sensing data', *Advances in Space Research* **33,** 268–273.
[35] Liu, J.Y. et al., 2001, 'Variations of ionospheric total electron content during the Chi-Chi earthquake', *Geophysical Resesearch Letters* **28**, 1383–1386.
[36] Lognonné, P. et al., 2006, 'Seismic waves in the ionosphere', *Europhysics News* **37**, 11–14.
[37] Parrot, M. et al., 2006, 'Examples of unusual ionospheric observations made by the DEMETER satellite over seismic regions', *Physics and Chemistry of the Earth* **31**, 486–495.
[38] Ward, S., 2002, 'Slip-sliding away', *Nature* **415**, 973–974.
[39] McGuire, B., 2005, 'Swept away', *New Scientist* **2522**, 22 October 2005, 38–41.
[40] Geist, E.L. et al., 2006, 'Waves of change', *Scientific American*, January 2006, 42–49.
[41] Titov, V. et al., 2005, 'The Global Reach of the 26 December 2004 Sumatra Tsunami', *Science* **309**, 2045–2048.
[42] Satake, K. and Atwater, B.F., 2007,'Long-term perspectives on giant earthquakes and tsunamis at subduction zones', *Annual Review of the Earth & Planetary Science* **35**, 349–374.

[43] Mills, E., 2005, 'Insurance in a climate of change', *Science* **309**, 1040–1044.
[44] Bengtsson, L., 2007, 'Tropical cyclones in a warmer climate', *WMO Bulletin* **56**, 1–7.
[45] Witze, A., 2005, 'Bad weather ahead', *Nature* **441**, 564–566.
[46] Donnelly, J.P. and Woodruff, J.D., 2007, 'Intense hurricane activity over the past 5000 years controlled by El Niño and the West African monsoon', *Nature* **447**, 465–468.
[47] Nyberg, J. et al., 2007, 'Low Atlantic hurricane activity in the 1970s and 1980s compared to the past 270 years', *Nature* **447**, 698–701.
[48] Mooney, C., 2007, *Storm world: hurricanes, politics and the battle over global warming*, Harcourt, Ed., p. 392.
[49] Reichhardt, T. et al., 2005, 'After the flood', *Nature* **437**, 174–176.
[50] McKenzie, J.A., 1999, 'From desert to deluge in the Mediterranean', *Nature* **440**, 613–614.
]51] Gupta, S., 2007, 'How a map of the English Channel explained Britain's island status', *Nature* **448,** xv.
]52] Ryan, W. and Pitman, W., 1998, *Noah's Flood: The New Scientific Discoveries about the Event that Changed History*, Simon & Schuster, p. 319**.**
[53] Laurance, W.F., 2007, 'Forests and floods', *Nature* **449**, 409–410.
]54] The town of Nouakchott saw its population changing from 20,000 in 1960 to 350,000 in 1987, creating major problems of adaptation for the management of the city (see Engelbert, P., 2001, *Dangerous Planet, Avalanche to Earthquake*, Tome 1, Sadinski, D. (Ed.), UXL, p. 446).
]55] Schiermeir, Q., 2005, 'The chaos to come', *Nature* **438**, 906.

5 The Changing Climate

I am far from supposing that the climate has not changed since the period when those animals lived, which now lie burried in the ice.
Charles Darwin, *The Voyage of the Beagle*

5.1 Miscellaneous evidence of climate change

A century ago, captain Larsen, commanding a whaling vessel, discovered fossil wood on the Antarctic Peninsula, where nowadays only a few lichens grow [1]. Apparently, the climate in the remote past was much warmer than today. Recently, some crocodile-like fossils were found near northern Greenland – animals restricted to subtropical parts of the world. These fossils are some 90 million years old, dating back to the mid-Cretaceous [2]. At that time mean annual temperatures at polar latitudes were apparently above 14°C. After this high point a cooling trend set in, and by 30–40 million years ago an ice cap may have formed over the Antarctic continent which gradually increased in extent and is still there today [3]. Of course, one has to take into account that continents have moved over geological timescales (section 2.4), but these locations have remained at high latitudes over the last 100 million years.

At the termination of glaciers one finds much clay-like and rocky material (the 'end moraine') which has been transported on the surface of the glacier or scraped from its bottom. The former may include large boulders, and the latter, stones with characteristic striations. In the Alps such material could be found far from the end points of glaciers, indicating a larger extent in the past. When around the middle of the 19th century the Swiss naturalist Louis Agassiz found similar material in the British Isles and elsewhere in northern Europe, he concluded that ice caps had covered large stretches of Europe [4]. Somewhat later the same was found to be the case in North America with the ice extending into Illinois and Kansas. Apparently, a cold period had gripped much of the northern hemisphere. Subsequently it was found that several such cold periods – glacials – had occurred interspersed with short warmer epochs – the interglacials. It also seems that other parts of the Earth were cooler during the glacials. Some controversy still exists on the amount of cooling in the tropics, but figures around 3–4°C for the sea-surface temperature have been proposed [5]. Global mean temperature has been estimated as nearly 6°C colder than present (pre-industrial) values [6]. The ice caps were very thick – several kilometers. With so much water locked up on the continents, the sea level was lower than now by up

to 120 meters. On the Greenland Ice Sheet, temperatures were some 20°C colder than today, and around 11,000 years ago the last glacial period came to a rather abrupt end. It may be no accident that soon thereafter agriculture began in the Middle East and in China.

The drop in sea level added dry land to the continents with far-reaching consequences. A land bridge appeared in the Bering Sea which allowed Asian tribes to cross into the Americas. It has been suggested that at the time northern America was covered by two ice caps, one in the northeast and one in the northwest, with an ice-free zone in between. Through this corridor the early migrants may have marched southwards some 12,000 years ago; it must have been a cold trip! In less than 2,000 years they reached the southern tip of South America. At the same time the climate became warmer and the sea level rose, destroying much of the evidence left by the migrants under the waters of the Bering Sea. Many big mammals (mammoth, mastodon, etc.) became extinct, perhaps because of the changing climate or, more probably, by hunting by the rapidly increasing human population (see Section 2.6.4).

More recently a favorable climatic period made the voyages of the Norsemen to Greenland, Vinland (New Foundland?) and other places possible. A regular traffic from Iceland to the western settlement on Greenland at 64°N had been established by AD 1,000, and several thousand Norsemen had their farms in two settlements. They explored far away, and on a cairn at latitude 72°55′ on the Greenland coast is written 'Erling Sighvatsson and Bjarni Thordarson and Eindridi Jonsson on the Saturday before the minor Rogation Day [25 April] built these cairns' [7]. Conditions in the far north must have been free of ice remarkably early in the season. Afterwards conditions deteriorated; the eskimos, who had moved north when climate improved, returned to the south. In a letter (1492) from the Pope, 'It is said that Greenland is an island near the edge of the world ... Because of the ice that surrounds the island sailings there are rare, for land can only be made in August when the ice has receded' [8]. A century later Breughel and others painted their winter landscapes in Flanders with lakes frozen over and people on skates (Figure 5.1). The 'Medieval Warm Period' had been succeeded by the 'Little Ice Age' at least in the North Atlantic region.

From the beginning of the 20th century the situation has changed and rapid warming has set in which intensified as the year 2000 approached. In the far north at Nenana on the Yukon some railroad engineers in 1917 placed a tripod on a frozen river and took bets on the moment in spring that it would fall through the ice. Because of the important sums of money involved, a careful watch was kept. Eighty years later this annual event still continues, but it occurred 5–6 days earlier in spring indicating significant warming [9]. Records of freeze-up and break-up dates of 26 northern rivers and lakes around the world show similar results [10]. In other northern regions the permafrost is melting, leading to the collapse of buildings. In Switzerland and in the Rocky Mountains glaciers have been withdrawing at a rapid rate – on average 10 meters per year [11], but in some cases much faster (Figure 5.2). In tropical Africa at the beginning of the 20th century Mt Kilimanjaro (5,840 meters) was still covered by



Figure 5.1 During the Little Ice Age in the 17th and 18th centuries painters frequently viewed scenes on ice; this painting by Hendrick Avercamp is at the Rijksmuseum in Amsterdam. (© Rijksmuseum Amsterdam.)

a characteristic ice cap 12 km^2 in area. Only 2 km^2 are left, and in 20 years tourists need no longer flock there to see the grandiose spectacle [12]. More importantly, the glacier-fed streams will no longer provide water to the inhabitants of the surrounding area during the dry season. Such changes in glaciers may also be caused by important droughts. An earlier reduction of the Kilimanjaro ice cap some 4,000 years ago was related to a three-centuries-long drought that caused much havoc to human societies in Africa and the Middle East. At that time a significant part of the ice still remained. Tropical glaciers in the Andes are now also rapidly disappearing [13]. And the Himalayan glaciers are retreating rapidly, risking first the catastrophic overflow of glacial lakes and, if trends continue, dry season water shortage in the Ganges valley where 500 million people live.

Still further south in Antarctica, some of the ice shelves have been disintegrating around the Antarctic Peninsula due to rapid local warming with a loss of more than 13,000 km^2 of floating ice (Figure 5.3) [14]. Future observations should show if this is a very local phenomenon or not. Since in some places the ice shelves contribute to the stability of the ice cap on western Antarctica, such developments are a source of concern. If the whole western ice cap were to melt, the sea level around the world would rise 6 meters, flooding large areas.



Figure 5.2 Retreat of the Muir glacier in Alaska. Over 63 years the glacier retreated some 12 km. Trees now grow in the foreground where, in 1941, there was only bare rock scraped clean by the glacier. (*Courtesy*: USGS National Snow and Ice Data Center, W.O. Field, B.F. Molina.)

So there is evidence of rapid climatic variations on timescales of decades, centuries and millennia. Some of these have been of regional importance only, but others are world wide. While the historal evidence indicates that such variations may have had a significant human impact, a much more quantitative approach is needed to find the causes and to determine what the future may hold.

5.2 The global climate system

The surface of the Earth is heated by the Sun, the heat flow from the Earth's interior being negligible in comparison. At the same time, the Earth radiates heat into space. The temperature on Earth is such that the heat absorbed from the Sun is balanced by the heat radiated into space. Not all of the incoming solar radiation is absorbed by the Earth, and some is reflected back into space. Even a dark soil would reflect several percent, but clouds and snow have a very much higher reflectivity. Averaged over the whole Earth, about 30% of the incoming radiation is reflected. The 70% that is absorbed is ultimately reradiated into space, not so much by the Earth's surface as by its atmosphere. It is here that the natural 'greenhouse' effect comes in (see Box 5.1).

The solar radiation is distributed very non-uniformly over the surface of the Earth. Per square meter, more comes in towards the equator than towards the poles. Reflection losses are low over areas with dark soils or vegetation and high over regions with much cloud or ice cover. This non-uniform heating leads to differences in temperature and pressure, and the latter drive the winds. About 1% of the solar energy is transformed into wind energy, but, ultimately, friction transforms it back into heat energy.

At a given pressure warm air is less dense than cold air. Thus, warm air tends to rise and cold air to sink. The rising of warm air in the tropics tends to set up a circulation where, at higher altitudes, the air flows poleward and near the surface in the opposite direction, but the rotation of the Earth deflects these flows and

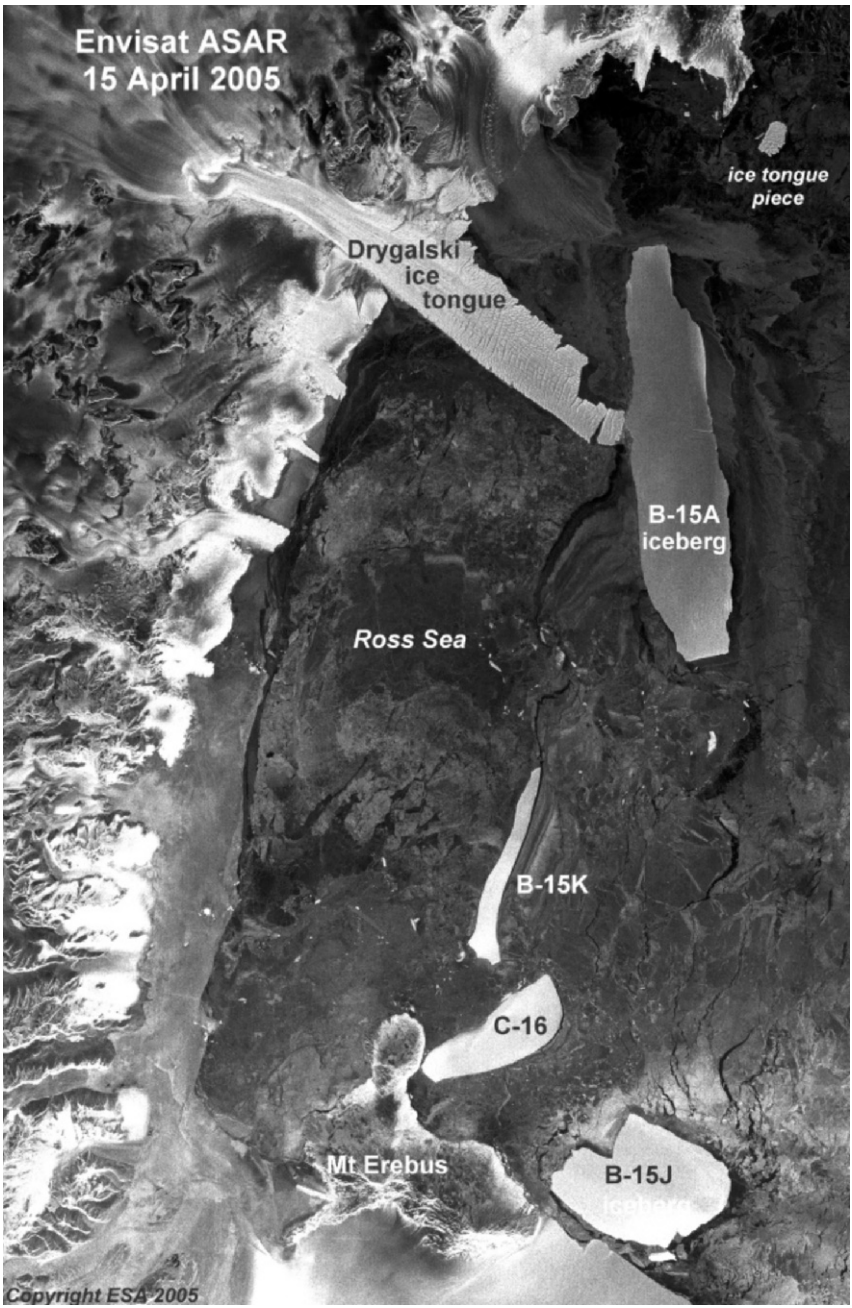


Figure 5.3 Floating sheets of ice colliding in Antarctica as observed by the satellite Envisat. Similar slabs of ice have resulted from the disintegration of the Larsen B ice shelf. When the icebergs drift into warmer waters they will melt rapidly. (*Source*: ESA.)

makes the circulation much more complex. The air at higher altitudes already comes down in the subtropics which, as a result, tend to be rather dry. The further poleward circulation of heat and humidity is affected by turbulent eddies, the mid-latitude storms.

Differences in temperature and salinity and the effects of the winds drive a circulation in the oceans. Although it is extremely slow owing to the great mass of the oceans, this circulation plays a vital role in transporting heat over the surface of the Earth. As part of the circulation, the 'Gulf stream' flows from the Caribbean towards Spitzbergen and contributes to making the climate in northwestern Europe relatively mild. During its northward flow the water cools. In the Arctic the cold salty water sinks and deep down it flows back. Perhaps not yet fully understood, this flow forms an integral part of a still larger 'thermohaline (salt) circulation' involving much of the Earth's oceans (Figure 5.4), and the atmosphere and the oceans thus form a coupled system. Constructing models of that system is a complex matter, although much progress has been made in recent years due to the availability of large computers, but some of the parameters entering these models remain rather uncertain. In the models there are two important concepts: feedbacks and forcings. If we make a small change to the system and if that change tends to amplify itself, we have a positive feedback with negative one in the opposite case. Radiative forcings are external changes influencing the system: If the Sun's luminosity were to increase or if the CO_2 concentration were to augment, it is a positive forcing; if volcanic eruptions put more dust into the atmosphere, the forcing is negative. Some examples of important feedbacks are given below.

Suppose that for some reason the high Arctic becomes somewhat warmer, and that some of the ice there melts. As there is little salt in the ice, a layer of fresh water will be formed. However, as fresh water has a lower density than salt water, it cannot sink, and the formation of cold bottom water may then be diminished. This may slow down the thermohaline circulation and thereby cool the northern regions, since less heat from the tropics will be transported northward. The initial warming may therefore be changed to cooling – a negative feedback – which will stabilize the climate by stopping the formation of melt water, unless the cooling is overdone, leading to a more vigorous circulation and an even stronger heating. In such a way, the system could oscillate between periods that are warmer and colder than average.

Suppose next that the cooling leads to an increase of the area covered by ice and snow. Since the snow has a high reflectivity, less solar energy is available to heat the surface and, therefore, additional cooling will occur. Here we have a positive feedback in which the cooling amplifies itself. In the same way, some initial heating may reduce the surface area that has snow cover, and increase the solar energy available for heating the surface. While this additional heating could destabilize the climate, this may be limited by other effects.

Dust blown from the land or transported by rivers may fertilize the oceans and foster the growth of plankton. Some of the small shelly animals that live on this may absorb CO_2 into their calcareous shells which ultimately form limestone.

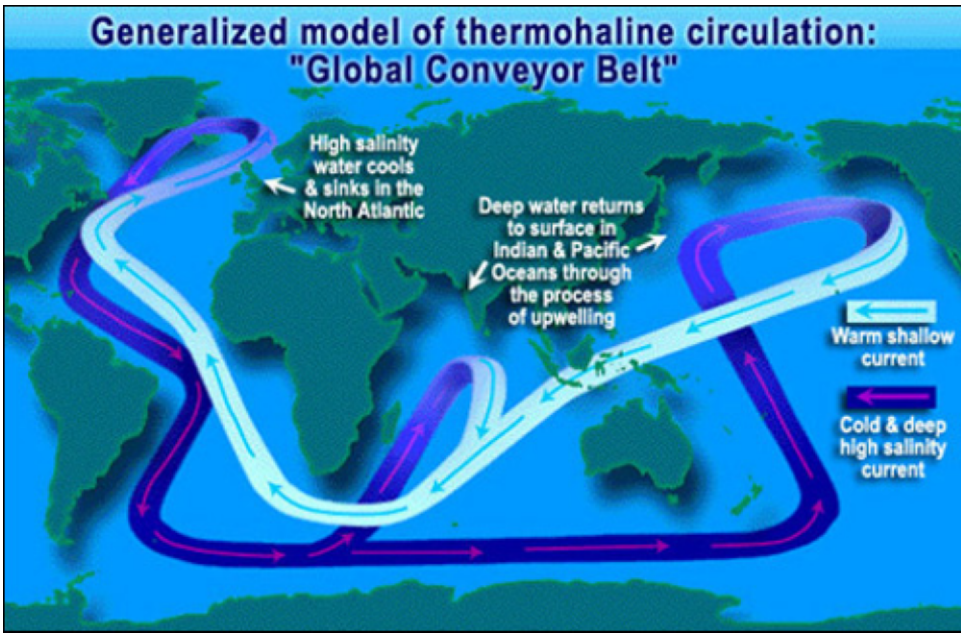


Figure 5.4 The thermohaline circulation. Though the ocean flow is slow, the total amount is more than 10 times larger than the flow through all the Earth's rivers. In the north Atlantic the surface flow of water from the equator warms the northern regions with more than 1,000 terawatts of heat energy equivalent to 100 times the present-day anthropogenic world energy production. (*Source*: NASA.)

This may diminish CO_2 in the atmosphere and lead to cooling by a reduced greenhouse effect. If the cooling reduces vegetation, more dust may be moved. On the other hand, weathering is encouraged by warmth, and so, in the long term, there might be less dust available in a cooler climate, drawing down less CO_2, increasing it in the atmosphere and thus counteracting the cooling.

If the ocean water warms, more will be evaporated and the concentration of water vapor in the atmosphere may increase. But as water vapor is a greenhouse gas, this could reinforce the warming. Thus, there are many processes affecting ice cover – oceanic circulation, CO_2 and water vapor content, to name just a few – that influence the climate and operate on different timescales. The general result tends to be a system where, overall, there is a certain balance, but where there are many possibilities for oscillations around the mean. Some have described this interplay of physical, chemical and biological factors as an almost mystical Gaia phenomenon. However, it is nothing more than a natural system following the laws of nature. On Earth the different processes have achieved an equilibrium that also permitted the biological factors to play a major role, but on both Venus and Mars this was not possible (see Chapter 9).

5.3 Climates in the distant past

In the context of the Earth's history the climate variations of the last millennia are nothing exceptional. Among the rocks more than 500 million years old we find evidence of early ice ages, while over 4,000-million-year-old sediments provide evidence for liquid water. This is actually quite remarkable. The early Sun was about 25% dimmer than at present, and the Earth's climate could have been expected to have been cooler, well below freezing on average. That such cold was avoided was probably due to a high abundance of gases such as CO_2 and methane in the atmosphere at that time, and recent and anticipated future warming has given CO_2 and the greenhouse effect a bad name. However, without the greenhouse gases (see Box 5.1) the Earth would be frozen over, as was demonstrated by Svante Arrhenius in 1896 [15]. In fact, the liveability of the Earth depends on there being more or less the 'right' amount of greenhouse gases – not too much and not too little – and preferably not with changes that are too abrupt. The biological world has much more adaptability for slow changes.

Box 5.1 The greenhouse effect

If the Earth had no atmosphere, its temperature would be such that the amount of energy obtained from the Sun would be exactly equal to that radiated by the Earth. Because the Sun has an effective temperature of nearly 6,000 K, it radiates most of its energy in visible light. With the Earth having a temperature near 300 K, it radiates in the infrared at 20 times longer wavelengths. The atmosphere is rather transparent to visible light, but more opaque to infrared radiation. As a result, that radiation is largely trapped in the lower atmosphere and will slowly diffuse to the higher parts from where it finally can be radiated away. But the temperature there is lower than at the surface, since otherwise there would be no upwards flow of the infrared radiation. Since a certain temperature there is needed to radiate the solar energy away, the temperature at the surface will be higher than that. In fact, the average emission temperature of the infrared radiation should be 254 K (–19°C) to balance the solar input. As a consequence of the greenhouse effect, this corresponds to 287 K (+14°C) at the surface, about 33 K higher. Not all of this difference is due to CO_2. Other greenhouse gases such as methane (swamp gas, CH_4), nitric oxide (N_2O), ozone (O_3) and industrial fluoro- and chlorocarbons play a role. Water vapor (H_2O) is also an important contributor to the greenhouse effect. However, while the other gases have a more or less fixed abundance in the short term, the water vapor content results from evaporation which depends on temperature and so forms part of a complex system of feedback.

Of course, we know few details about the early climate on Earth. However, some 600–700 million years ago one or more widespread ice ages have left traces on continental areas which, at the time, were rather near the equator (see Section 2.4.2). Later, other periods with important ice coverage occurred: first, briefly towards the end of the Ordovician ($\sim$460 million years ago), and subsequently more extensively in the later Carboniferous and the early Permian (320–280 million years ago). The former has not yet been understood, since CO_2 abundances were relatively high. The latter occurred at a time when much of the Earth's continental mass was assembled in the supercontinent Gondwana centered on the South Pole. CO_2 abundance was quite low [16], perhaps as a consequence of the early dense vegetation in the tropical swamps, where the carbon became fixed as coal. In the subsequent Mesozoic era (251–65 million years ago) there is not much evidence for ice caps. Climate was relatively warm, also in the polar regions, and most estimates of CO_2 concentrations are in the 1,000–3,000 ppmv range – a typical 'greenhouse climate'. The $\delta^{18}O$ record (see Box 5.3 on page 164) shows that, thereafter, temperatures remained high through the 'Eocene Climatic Optimum' (55–50 million years), but that by 50 million years cooling got underway which, with ups and downs, continued to the present day [17]. Even deep ocean temperatures fell by some 12°C as evidenced by the magnesium/calcium ratios in shells. Somewhat uncertain estimates of CO_2 concentrations show a decline to 300 ppmv or below after the beginning of the Miocene (24 million years ago) [18], comparable to recent pre-industrial values. So the 'greenhouse climate' was replaced by the 'icehouse climate'. In fact, the $\delta^{18}O$ and Mg/Ca records show that massive ice sheets developed 34 million years ago over Antarctica with remarkable rapidity. In two steps of less than 40,000 years duration, separated by 200,000 years, a complete ice cap was built [19]. Antarctic ice diminished some 8 million years later, but the east Antarctic ice sheet was fully re-established by 14 million years ago. In the meantime some ice had also formed over the Arctic regions, which increased in extent from 3 million years onward (see Box 5.2). At that time began the quasi-regular succession of glacial periods separated by brief warmer interglacials. The Holocene, the period that began some 10,000 years ago, is the most recent interglacial, with the ice caps on Antarctica and Greenland remaining but with most of the continental ice on Eurasia and North America having melted. The duration of typical interglacials was no longer than the Holocene. Could a new glacial period be expected in the not too distant future (see Section 6.4)?

The high CO_2 concentration during the early Eocene may have been related to volcanism at the time that the Atlantic Ocean opened up. Other important events may also be related to continental drift. The northward motion of Australia and the westward motion of South America broke their connections to Antarctica and thereby isolated the latter continent at a somewhat uncertain

Box 5.2 El Niño and La Niña

Usually the equatorial west Pacific is warm, but further east upwelling of cold water causes much lower temperatures. In fact, most of the ocean is cold and there is only a shallow pool of warm water. The cold water may be brought to the surface by the stirring caused by the trade winds. However, from time to time the trade winds weaken and the east Pacific also warms. Such events tend to occur towards the end of the year, and this has given the name El Niño (the Christ child) to the phenomenon. It is characterized by warm temperatures (up to 2°C or even more) and therefore by humid conditions with much rain on the normally dry Peruvian coasts, which may last for the better part of a year. La Niña is the opposite phenomenon. Both are part of the El Niño Southern Oscillation (ENSO), a cycle of pressure variations in the equatorial Pacific leading to changes in the trade winds. The ENSO effects are felt around the world [20]: sometimes El Niño has been associated with failure of the Indian monsoon and drought in Indonesia. While the El Niño effects are self-limiting by creating waves in the ocean that lead to upwelling, the event that sets off the weakening of the trade winds is still not very clear.

During the early Pliocene (5–3 million years ago) there appears to have been a permanent El Niño condition which prevented colder water from reaching the surface [21]. This may have been the cause of the still relatively warm conditions (3-4°C above present) during that period, even though the CO_2 concentrations were about the same as today. When about 2–3 million years ago the warm water pool in the tropical oceans became shallower, La Niña conditions became more frequent, El Niño more sporadic, and the world arrived at ice age conditions. Of course, many other factors played a role. The question has been raised if the current increases could lead to a deepening of the warm water pool and restore the permanent El Niño state with higher temperatures.

moment not far from the time that the major ice cap appeared [22]. The circumpolar oceanic circulation with strong winds at the surface may have increased biological productivity and thereby drawn down CO_2 into the ocean as carbonate. The collision of India with the Asian continent led to the uplift of Tibet, which affected atmospheric circulation in an important way. Some 4-5 million years ago the Panama seaway closed, separating the Atlantic and Pacific Oceans which must have affected the oceanic circulation. These and other tectonic events must have influenced world climate and biology in important ways. While each of these have had their adherents as a dominant factor in particular climatic events, it is by now clear that the principal global cause of the transition to the icehouse climate has been the reduction in the CO_2 concentration. In addition, when CO_2 is low, climate seems to be more sensitive to other factors.

5.4 The recent ice ages

The ice in the polar caps is quite old. Every year a new layer of snow is added at the top. So if the ice cap is stationary, an equivalent amount of ice has to be removed each year. The ice flows slowly towards the coast, and on Antarctica the mass loss is in the form of icebergs and melting of the shelves. On Greenland, in addition, run-off and bottom melting are important. When an ice core is drilled, the individual annual snow accumulation layers can be recognized. Since the ice at greater depth is compressed by the weight of the overlying matter, a precise analysis becomes more difficult. In Greenland, ice cores have been obtained which, near the bottom, are 123,000 years old and on Antarctica 810,000 years. The difference is due to the much lower snowfall in the latter, only about 25 mm of H_2O per year at the 3,230-meter-high Dome C at 75°S latitude. Here the mean temperature is –54°C at the top but the ice is near melting at the bottom.

Such ice cores give much detailed information. Air bubbles enclosed in the ice tell us how much CO_2 and methane there was in the atmosphere. Radioactive beryllium (^{10}Be) gives information on the intensity of cosmic rays and thereby on solar activity at the time. The relative abundance of the oxygen isotopes ^{16}O and ^{18}O, expressed as $\delta^{18}O$, or the hydrogen isotopes ^{1}H and ^{2}H (deuterium) expressed as δD, in the ice relate to temperature. Dust in the ice is more abundant when strong winds blow, and volcanic ash layers may confirm the chronology of cores in different locations. Ice cores give rather complete, high-precision data, but they are restricted to a few polar or high-altitude areas. Oceanic sediments have a much wider distribution and their biological content is quite informative. Different organisms prosper at different temperatures, and the isotopes in their shells relate to the temperature and to the composition of the ocean water which, in turn, gives information about the amount of water locked up in ice caps (Box 5.3).

At Dome C in Antarctica the EPICA consortium (European Project for Ice Coring in Antarctica) has drilled the deepest ice core to date which has reached 810,000-year-old ice [23]. From gas bubbles in the ice the deuterium to hydrogen ratio has been obtained, which is largely a measure of the temperature. The total range in the EPICA core corresponds to about 12°C. Especially during the last half million years, a succession of brief warm intervals is seen, including the present Holocene period, separated by longer cold periods. The characteristic pattern of a sudden warming to interglacial conditions followed by a more gradual irregular decline into cold glacial conditions is repeating on average about once every 100,000 years. Similar results for the first 400,000 years were obtained for the French–Russian Vostok ice core at a distance of some 560 km, which gives confidence that local effects are minor. The CO_2 variations in the air are accompanying the temperature variations.

The record of temperature or ice volume is not strictly periodic. Nevertheless, certain quasi-periodicities can be recognized. If a mathematical analysis of the whole record is made in which all possible periods are considered and the 'power'

Box 5.3 Information from isotopic abundances

The chemical properties of elements are defined by the number of electrons in their atoms, which is equal to the number of positively charged protons in their nuclei. The weight of the atoms is determined by the number of protons and neutrons, the latter being uncharged. When two nuclei have the same number of protons but different numbers of neutrons, they are called isotopes. Thus, the nucleus of an oxygen atom has 8 protons and usually also 8 neutrons, for an atomic weight of 16, ^{16}O. There are two other isotopes with 9 or 10 neutrons and so with atomic weights of 17 (^{17}O) or 18 (^{18}O). The abundance of ^{18}O on Earth is only 0.2% of that of ^{16}O, while that of ^{17}O is still five times lower. These isotopes have the same overall chemical properties, but there are subtle physical differences. When water evaporates, there is in the vapor a very slight deficit of the heavier molecules. So the snow that falls on the polar caps has also less ^{18}O, and so does the resulting ice. If much water is locked up in the ice caps, the remaining ocean water is then slightly enriched in ^{18}O. So δ^{18}O (equal to $(^{18}O/^{16}O)/(^{18}O/^{16}O)_{standard}$ -1) is a measure of the ice volume. It can be determined by measuring the isotopic ratio in the shells of past marine biota. However, also in biological processes and in precipitation, a δ^{18}O is created which depends on temperature. Thus, in the absence of significant ice caps, the δ^{18}O is a measure of temperature; but when there is much continental ice, other measurements are needed to ascertain the relative importance of the two effects. One way to do this is to measure the ratio of magnesium to calcium in the shells which depends on temperature, but also on the Mg/Ca ratio in the sea water that has not necessarily had the same value in the past as now. Other isotopes have been used to gain temperature information – for example, the ratio of deuterium (hydrogen with a nucleus containing also a neutron in addition to the proton) to 'normal' hydrogen ^{1}H, which defines a δD in a manner analogous to δ^{18}O. Because of the many factors that affect the isotope ratios, only carefully validated results can be trusted. For the use of radioactive isotopes in age determinations see Section 2.1.

in the spectrum of periods is determined, three groups stand out: periods around 100,000, 41,000 and 23,000 years. Also a 400,000-year timescale is important. The Serbian scientist M. Milankovitch in 1941 provided an explanation [24].

When some cooling takes place at high northern latitudes, the increased snowfall may lead to the formation of year-round snowfields. But as snow is a good reflector, the Sun's heat will be reflected into space instead of heating the Earth's surface, and will tend to amplify the cold. In Antarctica this effect would be less important, because the ice cap is permanent and cannot grow much because the continent is already fully covered. So Milankovitch could make the case that the insolation around 65°N is particularly important. But this

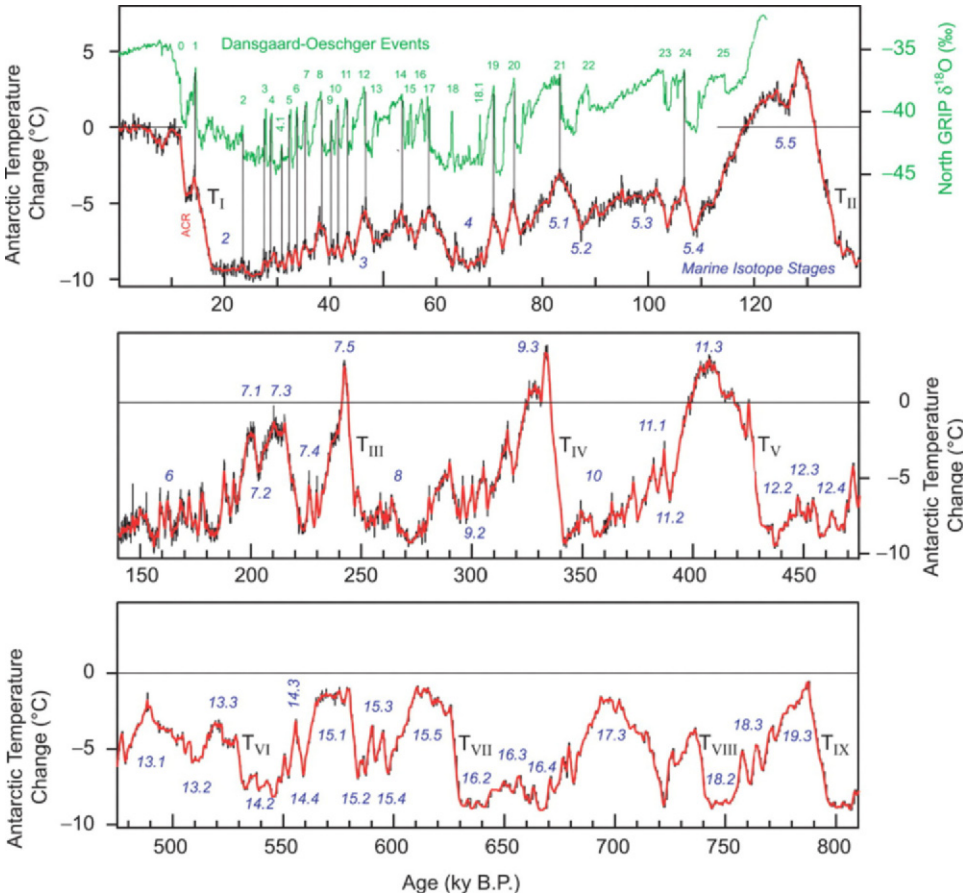


Figure 5.5 The EPICA results. In this reproduction of Figure 2 of reference [23] changes in temperature are shown during the last 800,000 years at Dome C in Antarctica. In the top panel, which has a magnified timescale, is also shown the sequence of events in Greenland, which shows much similarity in many details. The numbers in black correspond to the 'marine isotope stages' in sediments on the ocean bottom, while the T's refer to glacial terminations and the corresponding beginnings of warmer interglacials. During the first half of this period long interglacials alternated with cold periods, but interglacial warmth never reached present-day values. Beginning with the interglacial 400,000 years ago, the interglacial warmth increased but sometimes did not last very long; typically temperature increases towards the maximum were rapid, followed by a slower decline. The present interglacial, the Holocene in which we live, has not been particularly warm, though rather constant in temperature. At the resolution of this figure the warming of the last 80 years is not noticeable. (From J. Jouzel et al., 2007, 'Orbital and millennial Antarctic climate variability over the past 420,000 years', *Science* **317**, 793–796. Reprinted with permission from AAAS.)

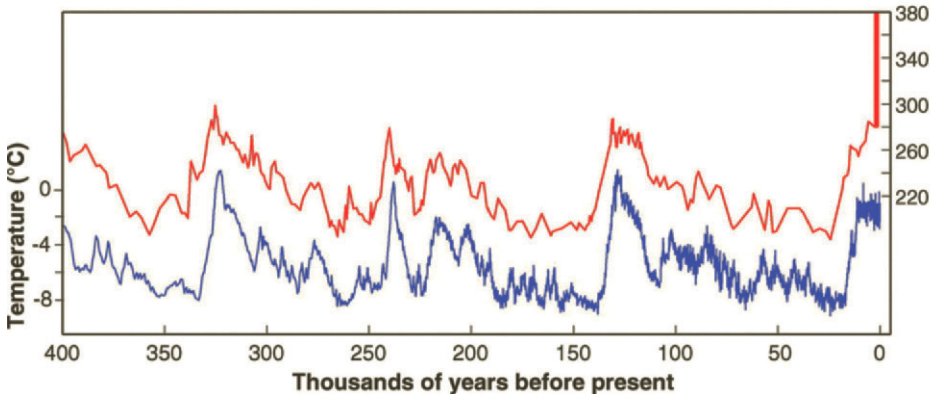


Figure 5.6 Temperature change and CO_2 concentration as derived from the Vostok ice core (Antarctica) results. The close correspondence between the two records is evident. Note the uniqueness of the high CO_2 concentration of 380 ppmv attained in the year 2006, which is mainly due to our burning of fossil fuels. (From A.V. Fedorov et al., 2006, 'The Pliocene paradox (mechanisms for a permanent El Niño)', *Science* **312**, 1485–1489. Reprinted with permission from AAAS.)

insolation is variable: because the orbit of the Earth is elliptical, it is sometimes closer to the Sun and at other times further away. The Earth's axis makes an angle with respect to the plane of its orbit – the so-called 'obliquity' – which is currently 23.5 degrees. The obliquity is responsible for the seasons; but the Earth's axis precesses (Figure 5.7). At present the situation is such that the Earth is closest to the Sun during the southern summer. Half a precession period later, i.e. in 10,500 years, this will be the case during the northern summer. Since the Earth receives more energy when it is closest to the Sun, the northern summers will then be particularly warm. During the northern winter it would be furthest from the Sun, but since it would in any case be dark at 65°N, this would not change things as much. The summer period would determine whether snow would melt. So it may be understandable that the precession period could influence the climate. But, in addition, the obliquity is not constant and varies with a period of 41,000 years with an amplitude of $\pm 1^{\circ}$. When the obliquity is largest, the precession effects are strongest. And finally the ellipticity – the eccentricity of the Earth's orbit – is variable with periods of about 100,000 and 400,000 years. Hence, the principal periodicities in the ice age climate could be understood as being due to astronomical effects – the properties of the Earth's orbit around the Sun.

So far, so good! But when we look in more detail, many problems appear. The effects of the variable eccentricity would be expected to be the smallest, while actually the 100,000-year periodicity is the strongest. The most remarkable feature is that the 100,000-year periodicity became dominant only some 900,000 years ago; before that the obliquity period dominated [25].

The current belief is that the astronomical periods certainly influence the

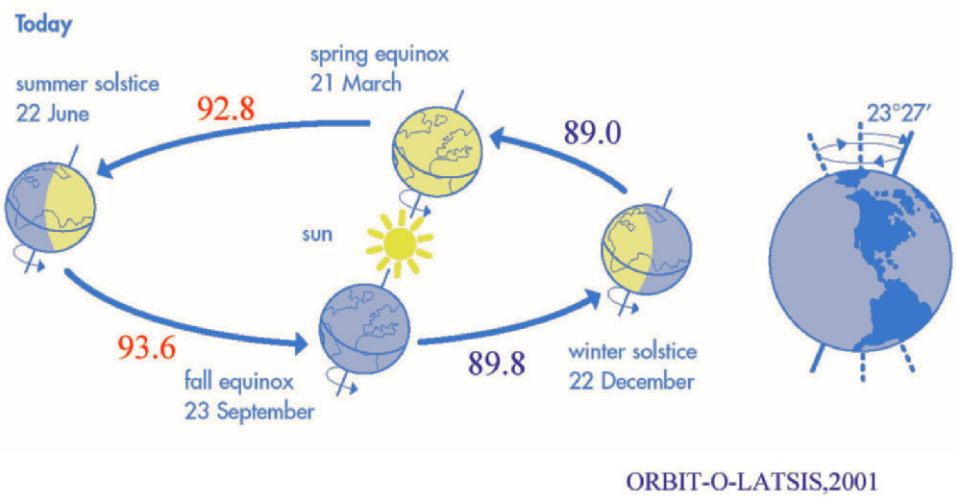


Figure 5.7 The present orbit of the Earth around the Sun. Because the orbit is elongated the Earth is most remote from the Sun around the summer solstice and so the northern hemisphere receives below average heat. But because of the precession of the obliquity the situation is reversed 10,500 years later to return to the present condition after 21,000 years. Thus, half that period later it is at the summer solstice that the Earth is closest to the Sun and receives above average insolation. In addition, the obliquity is variable with a period of 41,000 years, and at times of higher obliquity the seasonal effects become stronger. Finally the eccentricity, the oblateness of the orbit varies with periodicities of 100,000 and 400,000 years. At times when the orbit, is circular, the insolation is the same at both solstices and at maximum eccentricity the effects are strongest. All these periodicities are noticeable in the climate record of the ice ages. (This figure, courtesy M. Crucifix, was presented during the workshop on Solar Variability and Planetary Climates, held at the International Space Science Institute, Bern, Switzerland, June 6–10, 2005.)

climate, but that other effects are also important. For example, the CO_2 content of the atmosphere varies more or less in phase with the temperature. During the cold phases the CO_2 concentration tends to be around 180–200 ppmv, and during the interglacials, 280–300 ppm. There has been much discussion whether the CO_2 variations precede or follow the temperature variations, and whether the northern hemisphere warms up first, or the southern hemisphere, as seems to have been the case at the end of the last glacial period some 15,000 years ago. The real situation is that the climate system is very complex with many modes of variation that interact in a non-linear way but are synchronized to some extent by the orbital parameters.

What happens to the nearly 100 ppmv of CO_2 that disappear from the atmosphere during the glacial periods? It corresponds to some 250–300 gigatons of carbon, of the same order of magnitude as the amount of carbon in the form of CO_2 produced by the burning of fossil fuels. It has usually been thought that it

has found its way into the oceans. An interesting alternative is that it became locked up in the permafrost, the permanently frozen soil of the Arctic regions [26]. The steppes and tundra, where the mammoths roamed, covered huge areas with permafrost. Strong glacial winds spread fertile dust over these areas resulting in a gradually thickening layer with a high carbon content. In fact, in the Siberian permafrost today there is quite an amount of carbon (2–3%) from ancient grass roots, animal bones, etc. When at the end of the last glacial period the climate warmed and some of the frozen ground melted, the organic matter rotted away releasing the carbon into the atmosphere as CO_2. This was a fast self-accelerating process because the increasing CO_2 concentration caused further warming and melting of the permafrost in the areas where the temperature became high enough. So the CO_2 was slowly drawn down during the cold phase, while the increase during the warming could be very fast – the typical course of events as seen in Figure 5.6. The role of the remaining permafrost in future warming will be discussed in the next chapter.

The main conclusion from the record of the last million years is that for most of the time the climate was colder than at present. For the continents north of 40° latitude the mean temperature has been estimated as some 9°C below present values; the coldest extremes may even have been twice as cold as that [27]. Some brief interglacials separated the cold phases, with the current interglacial, the 'Holocene', having started around 10,000 years ago. Holocene temperatures generally have been rather stable. It is important to realize that such stability has been the exception and that some of the warm periods had a shorter duration than the Holocene to date. We know, of course, much more about the last glacial period and its ending than about earlier times, so we shall consider this last period in some more detail.

Around 120,000 years ago, the previous interglacial had come to its end. An unstable slowly declining phase followed, until 21,000 years ago when the Last Glacial Maximum was reached with temperatures in Greenland more than 20°C below present values. The instabilities during the glacial period were very strong and have been felt in many parts of the world. The temperature would rise rapidly (by perhaps 10°C) and, after a millennium, just as quickly return to the previous value (the so-called Dansgaard–Oeschger events). This suggests that as the climate cooled, different equilibrium states were possible and that minor disturbances could tip the climate from one to the other. The sudden warming was sometimes accompanied by a prolific calving of icebergs (Heinrich events) which transported the heavy rocks now found in the ocean in well-defined sedimentary layers. The melt water from the icebergs may have stopped the thermohaline circulation and caused sudden cooling.

By the time the Last Glacial Maximum was reached, large parts of North America and Eurasia were covered by ice sheets more than 1 km thick, and an early observer might have wondered how one could ever return to more clement conditions. But then, 14,600 years ago, the temperature briefly shot up to reach values not much less than those today (Figure 5.8). However, soon temperatures declined just as rapidly and during the 'Younger Dryas' for about 1,000 years

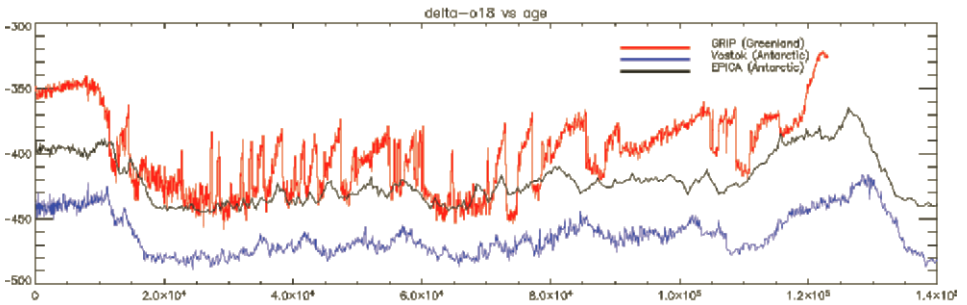


Figure 5.8 The $\delta^{18}O$ record over the last 40,000 years at central Greenland, GRIP, at Dome C in Antarctica, EPICA, and at Vostok some 560 km away. The $\delta^{18}O$ values are a rough measure of temperature with the range in Antarctica corresponding to a range of 9°C and at the GRIP site of slightly over 20°C. The sharp peaks in the GRIP curve are the Dansgaard–Oeschger sudden warming events (see text). The last of these, at 14,500 years ago, reached almost present-day temperatures, but was followed by the very cold Younger Dryas (named after the reappearance of the Arctic plant *Dryas octopetala* in southern Scandinavia). The temperature rise signaling the end of the ice age began some 17,000 to 18,000 years ago in Antarctica. It was followed around 13,000 years ago by the Antarctic Cold Reversal that was much less severe than the Younger Dryas in the north, which occurred slightly later. The temperature rise since AD 1920 is not visible on the scale of this image. (*Source*: Wikipedia.)

deep glacial conditions were re-established. The next warming, 11,500 years ago, was more successful. The temperatures went up by not much less than 10°C [28] within a few decades, with wind-blown dust diminishing in a time interval of no more than a few years. Apparently a sudden switch in the climatic system took place, and the previous conditions were never re-established until the present day. It is probable that the 'Younger Dryas' was produced by the catastrophic emptying of a large lake, Lake Agassiz, into the North Atlantic. This put a fresh water lid on the region where deep water formation takes place, and stopped the thermohaline circulation and the transport of warmer water to the North Atlantic.

Following the Younger Dryas the temperature continued to rise and appears to have been even one or two degrees warmer than today in the early Holocene. The most noteworthy event has been a brief cold snap 8,200 years ago (~2–3°C ?). It may have been related to another pulse of melt water. It is interesting that the beginning of the Holocene is more or less coincident with the early appearance of agriculture.

During the Last Glacial Maximum a large amount of water was locked up in the ice caps. As a result, the sea level was 120 meters lower than today. The rapid increase in temperature led to an equally rapid rise in sea level which, at times, reached values of more than 1 meter per 25 years [29]. Still the final melting of the continental ice sheets took time, and it was not until about 6,000 years ago that the last remnants of the North American ice cap had vanished. At the time

of the first beginnings of the Egyptian civilization, Quebec was still covered by ice!

The last interglacial period (the Eemian), from around 130,000 to 120,000 years ago, was significantly warmer than the Holocene, reaching a peak of perhaps 5°C above present-day values [27]. Also sea level was some 5 meters higher than now, showing that much ice had melted. However, the level of the top of the Greenland Ice Sheet (GIS) at the summit did not go down much. More detailed analysis suggests that perhaps 3 meters of water came from the GIS [30]. An additional amount may have come from the West Antarctic Ice Sheet. The GIS would have been a very steep ice dome with part of the more coastal areas ice free.

The last interglacial lasted no longer than the Holocene. So should we expect the Holocene to come to an end and be followed by a new ice age? Before arriving at such a conclusion we should look at the Earth's orbit again. From Figure 6.2 it is seen that the Eemian interglacial was associated with an exceptionally high summer insolation and was terminated by a very deep minimum. The Holocene began with a somewhat more modest insolation, which will not vary much during the coming 45,000 years because the eccentricity of the Earth's orbit has become negligible. With a circular orbit the insolation forcing becomes very weak. The last time that a similar situation prevailed was during the interglacial 400,000 years ago. The EPICA ice core gives detailed information about that interglacial which seems to have had a longer

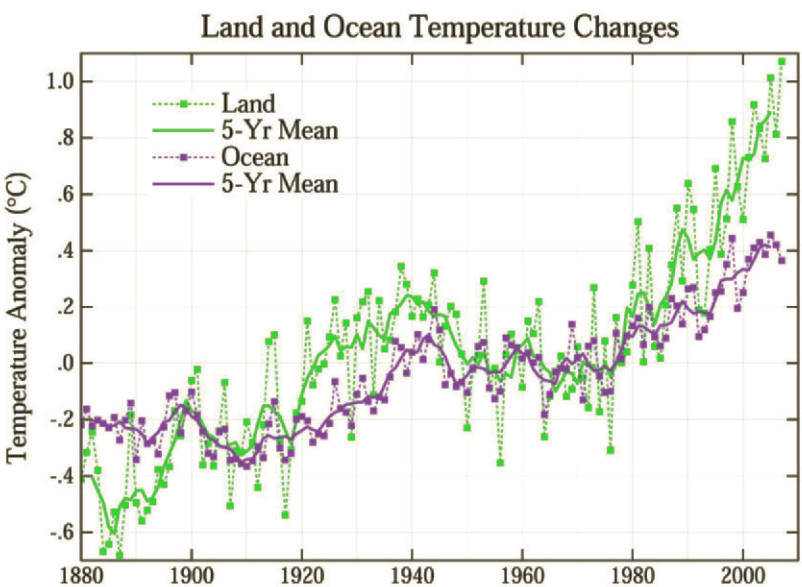


Figure 5.9 Global temperatures over land and over the oceans with respect to the mean temperature for 1951–1980. (*Source*: NASA–GISS.)

duration (Figure 5.5). In fact, if we fit the Holocene temperature curve to that of the interglacial 400,000 years ago, we would conclude that the Holocene might still continue to be warm for another 15,000 years. However, since the orbital configuration 400,000 years ago was not exactly the same, this curve-fitting remains somewhat uncertain. In any case, future developments may be very different because the anthropogenic production of greenhouse gases is leading to a very different atmospheric situation. We shall return to this in Chapter 6.

5.5 Recent climate

During the last 150 years thermometric records are available from many places on Earth which allow us to estimate its global mean temperature. Care is needed, however, to exclude data from the 'heat islands' around big cities. It should be noted that southern hemisphere data are quite incomplete. Over the last 30 years Earth observation satellites have obtained accurate global data of land and sea surface temperatures, and for different layers of the atmosphere. Early controversies on the interpretation of the satellite data appear to have been resolved. In addition, winds, precipitation and other parameters of the atmosphere and oceans are regularly measured with a variety of satellites. It is particularly important to ensure the continuity of the data by having an overlap between different satellites; if not, any apparent changes may not be convincing owing to the risk of slightly different sensitivities in subsequent generations of satellites.

Inspecting the record (Figure 5.9) we see that globally averaged temperatures were relatively stable until AD 1920. Thereafter warming set in until the early 1940s which resumed with even greater vigor around 1975 to the present. Total warming of the globe has amounted to about 0.8°C over the 1880–1920 average. Over land the increase was about 50% larger, with the large thermal inertia holding back the increase over the oceans. In the Arctic the rise was particularly spectacular (Figure 5.10). On lands south of the equator, warming was on average much less. Some parts of Antarctica seem to have cooled, although around the Antarctic peninsula warming was strong and probably contributed to the disintegration of some ice shelves.

In 1989, after the record warmth of 1988, climatologist James Hanssen made a bet that one of the first three years of the 1990s would be even warmer [31]. At the end of 1990 he collected his bet. Not only that, but out of the last 10 years nine have exceeded the global temperature of 1988. The warmest years, 1998 and 2005, exceeded 1988 by more than 0.4°C.

The thermometric record becomes more and more incomplete towards the middle of the 19th century and more indirect indicators have to be looked for. Tree-ring records are frequently used. Trees outside the tropics grow mainly during the summer, so each year the stem thickens by a new layer. By counting these layers we can date the trees. If conditions are favorable, relatively thick, dense layers may grow. It appears that the characteristics of the layer correlate

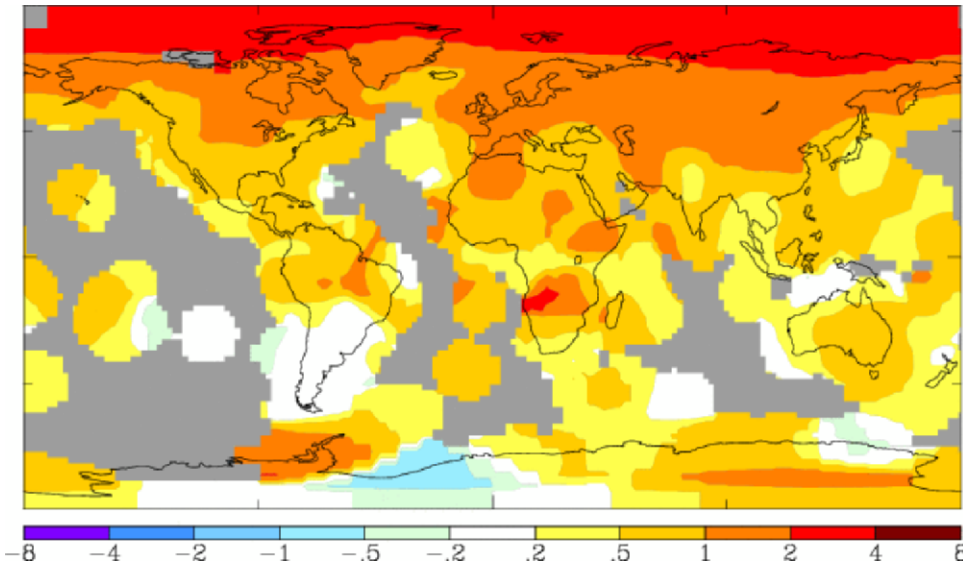


Figure 5.10 The difference of the temperatures (in °C) at the Earth's surface during 2005–2007 and that of the 1951–1980 mean. Grey areas indicate regions without data. Since the early data may not always be reliable, not every detail of the map need be significant. The strong warming in the Arctic is evident. (*Source*: NASA.)

well with temperature and with rainfall. During the last century and a half both tree-rings and thermometric records are available, and so we may calibrate the tree-ring characteristics and determine temperatures further in the past. In the same way corals may be used to obtain ocean surface temperatures. In addition, the length of glaciers, seeds in layered sediments in lakes, and shells in ocean sediments give further information. Also important are the 'borehole' data [32].

Suppose that at some moment the surface of the Earth warms. Heat conduction will then propagate some of that warmth to a greater depth, but because rock is a very poor conductor, it will take years before that 'heat wave' reaches a greater depth. In typical places it may take 500 years before a depth of 500 meters is reached, which implies that the temperature some 500 meters below the surface contains information about conditions half a millennium in the past. Of course, if the surface cools it is a 'cold wave' that propagates inward. These effects come on top of the general increase of temperature with depth owing to heat conduction from the hot interior, which should be unchanging on timescales of millennia. For studies of oil and other resources, many boreholes have been drilled all over the world where the temperatures can be measured. Although the boreholes give rather direct evidence of past temperatures, other problems may result from changes in snow and vegetation cover.

In Figure 5.11 [33–36] we have plotted the run of the mean temperature of the northern hemisphere for the last 1,800 years derived by two different sets of

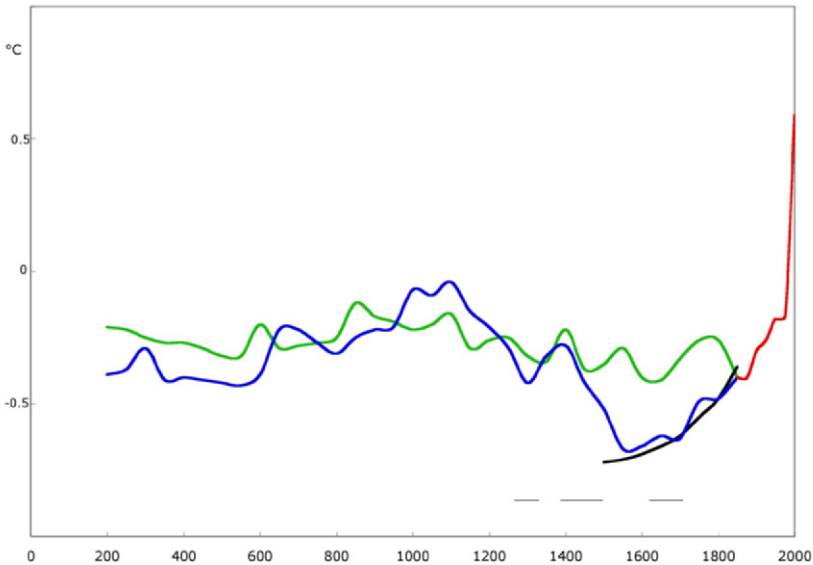


Figure 5.11 Annual mean northern hemisphere temperatures in °C with respect to the 1961–1990 average. The green curve is from P.D. Jones and M.E. Mann [33], the blue curve from A. Moberg et al. [34], both derived from calibrated proxies, and the black curve indicates the borehole data from H.N. Pollack and S. Huang [32]. The red curve represents the recent instrumental record from P.D. Jones [35]. The three bars lower down are periods of particularly low, measured or inferred Sunspot numbers from S.K. Solanki et al. [36].

authors. The first one by Jones and Mann, is based primarily on data with a high time resolution, like tree-rings [33]. The other, by Moberg et al., also includes less precisely dated century scale data, like those from lake or ocean sediments [34].

Moreover, it is based on different mathematical methods of analysis. For the last 500 years the latter shows almost twice the temperature variation as the former, which agrees better with the borehole data. For the moment it is not clear which curve is to be preferred.

The most remarkable aspect of Figure 5.11 is that, following two millennia of ups and downs with more modest amplitudes, the temperature shot up in two steps during the 20th century, reaching levels well above anything seen during the preceding 18 centuries. This suggests that something has changed rather suddenly in the climate system. As we shall see later, that something appears to be the increase of carbon dioxide and other greenhouse gases in the atmosphere.

Two features in the pre-industrial record (Figure 5.11) have been noted: the relatively high temperatures AD 900–1200, which have been baptized the 'Medieval Warm Period', and the 'Little Ice Age' with lower temperatures AD 1500–1800. Both these periods are recognizable in the North Atlantic area. However, it is not clear how widespread they were, and different areas have

somewhat different periods of maximum warmth or cold. Three types of causes have been proposed for the temperature variations of the last millennia: changes in the luminous output of the Sun, effects of aerosols from volcanic eruptions and anthropogenic greenhouse gases.

5.6 Changes in the Sun

Variations in the distance of Sun and Earth, as discussed by Milankovitch, are not the only possible sources of changing insolation. Real changes in the Sun's luminosity might also play a role. Variations in the solar radiation have an immediate effect on the Earth's atmosphere. With everything else remaining the same, a 1% increase in solar output would correspond to an increase in the global temperature of 0.7°C. But while the orbital variations at the basis of Milankovitch-type theories can be precisely predicted, the intrinsic solar variations are only partially understood.

Changes at the solar surface are evident. Sunspots – dark magnetic spots – come and go in an 11-year rhythm (Figure 5.12) [36]. It might have been expected that at sunspot maximum the Sun's radiation would be a bit weaker, since the spots radiate less, being cooler than the rest of the solar surface (4,000°C instead of 5,500°C). However, satellite measurements during the last three sunspot cycles have shown the opposite to be the case (Figure 5.13) [37]: at sunspot maximum the solar energy output increases by some 0.17%. The dimming in the spot is compensated by some extra energy related to the magnetic activity which apparently increases the total radiation (Figure 5.12).

For many years scientists had been looking for 11-year cycles in the climate, without much success. However, longer term variations also occur in the sunspot record, and recently these appear to have stronger effects. Although the Chinese had been recording large sunspots for the last two thousand years, more precise studies became possible only after the invention of the telescope. Galileo and others described many spots in 1610 and the following years, but after 1640 very few spots were seen. This 'Maunder minimum' lasted till around 1700, after which spots reappeared with their 11-year periodicity. It was subsequently noted that the Maunder minimum was a period of unusual cold in Europe – part of the 'Little Ice Age' – and so it seemed tempting to connect the two. But this correlation might also be just accidental, and therefore it is necessary to study a longer period. Unfortunately, adequate sunspot records do not go back further in time.

The Sun is surrounded by the corona, a hot (2,000,000°C) tenuous gas, which is heated by magnetic activity. The light of the corona is rather faint and can only be seen when the solar disk is hidden by the Moon during a solar eclipse, or nowadays also by satellites. The outer parts of the corona stream outwards which give rise to the 'solar wind' – a fast flow of gas (400–1,000 km) with magnetic fields, that pervades the Solar System. From time to time the activity at the solar surface leads to a more stormy flow which perturbs the magnetic field of the

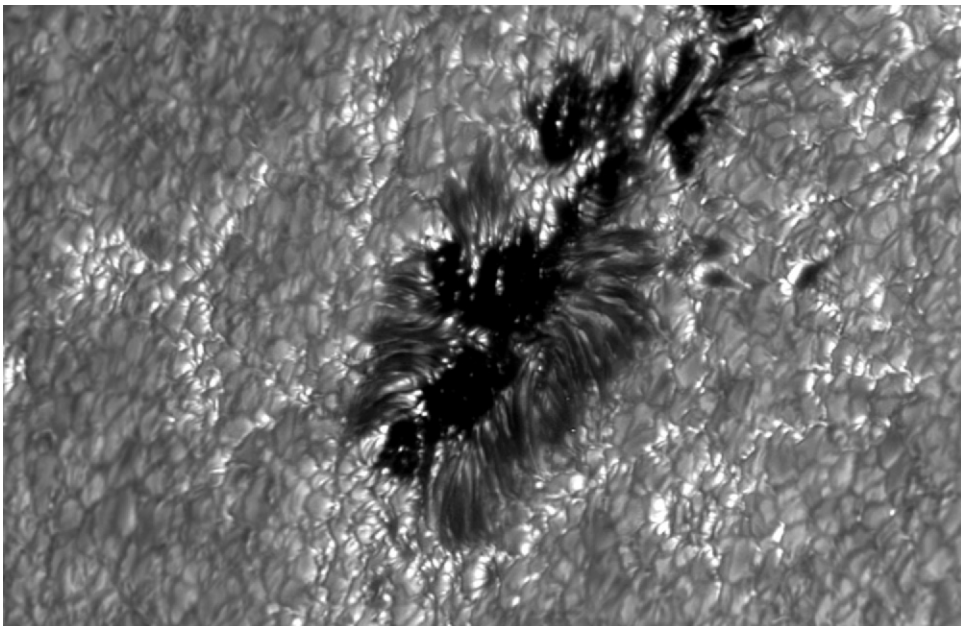


Figure 5.12 The solar surface. A dark sunspot reduces the solar luminosity, but bright magnetic 'faculae' increase it. As a consequence, the Sun is most luminous at sunspot maximum. Image taken with the Swedish Solar Telescope at La Palma. (*Source*: V. Zakharov and S.K. Solanki.)

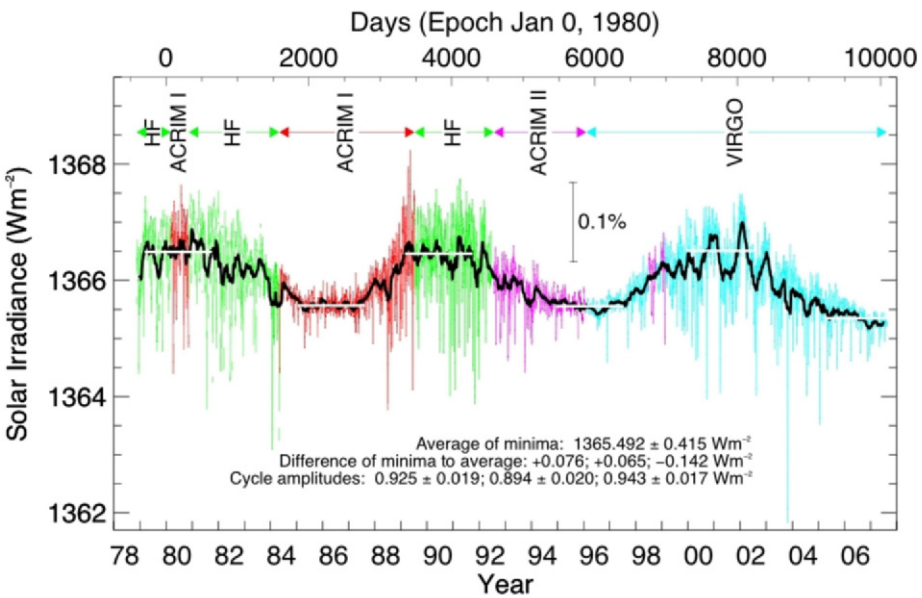


Figure 5.13 The solar irradiance in W/m^2 during three sunspot cycles. (*Source*: C. Fröhlich. This figure is updated from reference [37].)

Earth and allows energetic particles to reach the upper atmosphere where they excite the local gas and give rise to the Aurora Borealis – the 'northern lights' – and their southern counterparts. It is interesting that during the Maunder minimum there were very few reports of auroras, while also during solar eclipses no mention was made of the corona. It therefore seems that solar activity had come to a standstill. Since during the 11-year solar cycle lower activity corresponds to slightly lower solar luminosity, it would seem not implausible that during the Maunder minimum solar luminosity would have been particularly low, thereby cooling the Earth. However, our present understanding of possible changes in the solar luminosity suggests that these are quantitatively insufficient [38].

One method for determining solar activity beyond the historical record is based on ^{14}C, an isotope of carbon (see Box 2.1 on page 15). ^{14}C is produced as a result of cosmic rays striking nitrogen nuclei in the atmosphere. Cosmic rays are energetic particles filling our galaxy, as discussed in Chapter 3. On their way to the Earth they first meet the solar wind with its magnetic fields, which sweeps some away before they reach the atmosphere. Since the solar wind is stronger when there is much activity on the Sun, the flux of cosmic rays then is lowest, and therefore the ^{14}C production is also reduced. Thus, the ^{14}C production rate is diagnostic for solar activity. ^{12}C and ^{14}C have almost identical chemical characteristics; both make CO_2 molecules and these will be incorporated in trees and other organic matter. So if at some moment there are more ^{14}C atoms, this will be noticeable in tree-rings. Although there are many complications in practice, the ^{14}C record therefore corresponds to a record of solar activity. In fact, during the Maunder minimum the production of ^{14}C was relatively high. Two other periods during the last millennium have been found with similar characteristics: 1420–1540 and 1280–1340. Both correspond to relatively cool periods in the climate. Going back further in time, some scientists have concluded that during the last 10,000 years a climatic cycle of 1,500 years may also correlate with ^{14}C production, and therefore relate to solar luminosity variations [39]. The ^{14}C results involve the complexities of the carbon cycle. Perhaps ^{10}Be, which is also produced by cosmic rays, gives a more straightforward measure of solar variability without all the problems associated with the complex carbon cycle.

The more recent sunspot record after the Maunder minimum shows a tendency to increased activity. The magnetic field of the Sun directly measured during the last century also shows a substantial increase. Could it then be that the spectacular warming during the century was due to solar activity? Before answering the question we have to look at another influence on climate – volcanism.

5.7 Volcanic eruptions

In 1812 Mount Tambora on Sumbawa Island in Indonesia began rumbling. Three years later, in the evening of 10 April 1815, huge columns of 'fire' were seen to

rise from the mountain, explosions were heard at a distance of more than 1,000 km and the 4,300-meter-tall mountain was reduced to a caldera whose bottom is only 2,200 meters high [40]. This was probably the largest volcanic eruption of the last 10,000 years, and a huge amount of dust is believed to have been thrown into the stratosphere, where some of it stayed for a few years. In 1816 brilliant sunsets were reported in London and a sort of 'dry fog' over New England, indicative of dust high up in the atmosphere. Historically the year 1816 is known as 'the year without summer'. Snow fell in every summer month in New England, Europe was cold and shortfalls in agricultural production with famine resulted [40]. In 1600 the volcano Huyanaputina in Peru underwent one of the major eruptions of the past several centuries [41]. The summer of 1601 was very cold with freezing weather in Italy extending into July. From tree-rings it seems to have been the coldest summer of the last 600 years in the northern hemisphere. In 1991 Mt Pinatubo in the Philippines had a more modest but still large eruption, and in 1991/1992 a period of increasing global temperatures, culminating in the record year 1990, came to a sudden, though temporary halt [42].

It is now clear that the fine sulfate dust thrown into the stratosphere reduces the solar radiation which reaches the Earth's surface and is responsible for the cool years which follow major volcanic events. If the dust is only thrown into the troposphere – the lower part of the atmosphere – it rains out quickly and the effects are minimal. But some of the fine dust in the stratosphere may remain for a year or longer.

5.8 Anthropogenic CO_2

During the last several hundred million years, the continents were largely covered with forests or meadows. Part of the organic matter was later buried under subsequent layers of sediment, forming coal, petroleum and natural gas. But as volcanoes continued to exhale much CO_2 and other gases, an equilibrium was established with the buried organic carbon being replaced by fresh carbon compounds from the volcanoes. We are now burning much of the buried carbon at a rapid rate, liberating the CO_2 into the atmosphere and, as a result, atmospheric CO_2 is on the increase. Deforestation makes an additional contribution. Rice cultivation, cattle raising and industrial activities have increased the two other main greenhouse gases, methane (CH_4) and nitrous oxide (N_2O).

At the Mauna Loa Observatory in the Hawaiian Islands, far from local sources of pollution, the amount of CO_2 in the atmosphere has been measured since 1958. Starting at 315 ppm (parts per million), it has been increasing continuously to reach 380 ppm today (Figure 5.14) [43]. It is estimated that before the industrial age began it was around 275 ppm. Since CO_2 is the principal greenhouse gas in the atmosphere, it is to be expected that these increases should result in a warmer climate. How much warmer can only be established on the basis of detailed models that encompass all the complex interactions in the

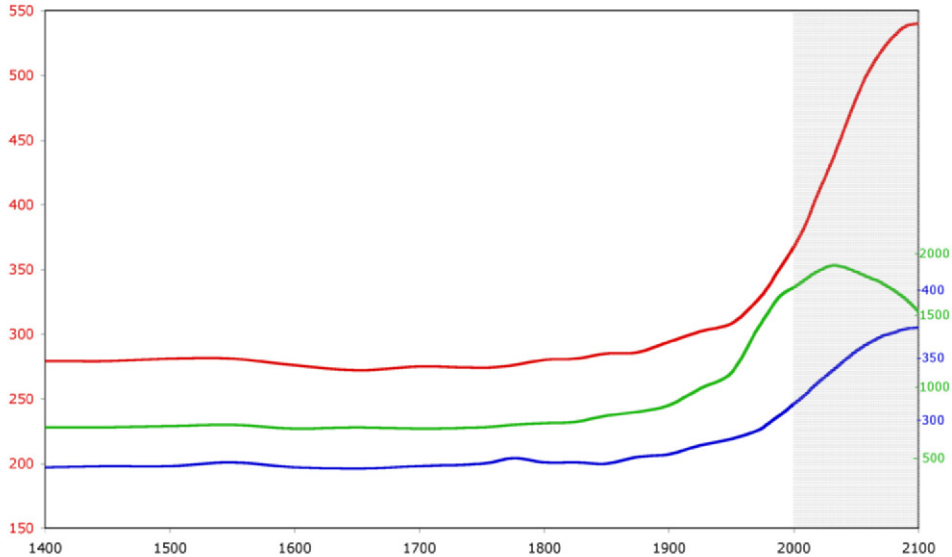


Figure 5.14 The concentrations of carbon dioxide (CO_2) in ppmv (parts per million by volume), methane (CH_4) in ppbv and nitrous oxide (N_2O) in ppbv from AD 1400 to 2000 and the projected changes after 2000 (in the gray area) on the B1 scenario (see Table 6.2 for the IPCC scenarios). Note that the B1 scenario, which leads to a doubling of the pre-industrial CO_2 concentration by 2100, is one of the more favorable scenarios with a relatively low input of anthropogenic CO_2. (Based on data from C. MacFarling Meure et al. and Tables II.2.1–II.2.3 in the IPCC Third Assessment Report, 2001 [43].)

ocean and in the atmosphere. Much progress has been made by model building, but much theoretical and experimental work remains to be done.

The amount of CO_2 produced annually is about double the increase in the atmosphere. Much of the rest appears to be absorbed by the oceans, which contain much more dissolved CO_2 than is present in the atmosphere, and also by the terrestrial biosphere. Increasing ocean temperatures would tend to reduce the capacity of the oceans to dissolve CO_2. As a result of the uncertainty in the impact of climatic warming on the oceanic CO_2 absorption, predictions of the future atmospheric CO_2 are somewhat uncertain. Also, vegetational changes could be important. But overall further increases are to be expected. Other greenhouse gases such as methane (CH_4) and nitrous oxide (N_2O) contribute additional warming.

5.9 Interpretation of the recent record

The observational record of the past two millennia (Figure 5.11) should in principle allow us to disentangle the important factors in the climate during that period. Of course, the limited accuracy in these 'observations', combined with

the uncertainty in the models, does not allow a very quantitative comparison, the more so since the amplitude of the solar luminosity variations is still a subject of debate. Nevertheless, the influence of the Sun is suggestive. The strong minima around AD 1300 and the Little Ice Age ($\sim$1500–1800) largely correspond to periods of low solar activity, while the low variability period with about average temperatures is also rather average as far as solar activity is concerned. The modest temperature increase in the first half of the 20th century may be partly related to the increase in solar activity combined with some increase in greenhouse gases. The absence of major volcanic eruptions during that time may also have played a role. The temperature increase after 1975 is seen to be extraordinary in the whole record and may be explained by the equally extraordinary increase in CO_2 concentrations to levels not seen for at least 740,000 years. In more detailed analyses of the pre-1900 temperature record, the conclusion has been that some 60% of the variance was due to solar and volcanic influences, but that these have been only of modest importance during the last three decades [44]. However, a climate model for pre-industrial Europe without any change in the Sun or volcanic effects, shows a variability very similar to that of the green curve in Figure 5.11, with the same 0.3°C fluctuations over periods of 50 years or so. Thus it remains difficult to disentangle the random variations due to natural internal variability from those due to external forcings [45].

There is an interesting side story to possible human influence on the climate. We have seen that the Earth emerged from the last glacial period at a time when the summer insolation in the north was high. Some 10,000–11,000 years ago maximum insolation was reached and began to decline, attaining a minimum at about the present time. Historical precedent thus suggests that several millennia ago a new glacial phase could have been initiated. Instead, the temperature remained more or less the same. It has been suggested that early deforestation caused a modest increase in CO_2 and early agriculture in methane just enough to avoid a new glacial phase [46]. Nonetheless, the very low solar activity combined with the near minimum insolation in the northern hemisphere caused quite low temperatures in the 17th and 18th centuries. As a result, the ice fields in northern Canada may have begun to form an ice cap. The resulting increased reflectivity of the land could well have tipped over the climate into a glacial decline [46]. Fortunately, perhaps owing to random fluctuations in the climate system, or owing to these small increases in CO_2 and methane, this did not happen. With the more rapid subsequent rise in these gases, such a risk now seems remote. Instead, excessive warming is expected.

5.10 The ozone hole

Sunlight dissociates atmospheric oxygen (O_2) molecules into oxygen atoms (O). Subsequently, one O may react with an O_2 to form ozone (O_3). O_3 may be dissociated by sunlight into $O + O_2$ or be destroyed by $O + O_3 \rightleftarrows 2O_2$. In the Earth's atmosphere these reactions lead to an equilibrium with typical ozone

abundances of a ppmv in the stratosphere. Although the O_3 abundance is low, the ozone layer is of fundamental importance to life on Earth because it shields us from the solar ultraviolet radiation. The integrated amount of O_3 through the atmosphere is measured in Dobson Units (1 DU = 2.7×10^{20} molecules per m^2) and before 1980 amounted from around 260 DU in the tropics to 280–440 DU at higher latitudes. Concerns were expressed at an earlier stage about the anthropogenic destruction of stratospheric ozone, in 1971 about nitric oxide (NO) from high-flying supersonic planes [47], and in 1974 from chlorine and bromine resulting from the industrial production of long-lived chlorofluorocarbons [48].

In the late winter and early spring of 1984, an unexpected reduction in ozone was discovered over Antarctica. From a long-time October average of 300 DU over the Halley station at 75°S, only 180 DU were left [49]. Since then satellite observations have shown that these reductions were continent wide and that, in attenuated form, even the southern parts of South America were involved. With much interannual variability the ozone hole worsened, and in October 2006 a record was set both for its extent and its depth with, in places, values below 100 DU [50]. At mid-latitudes small ozone reductions were measured, while the Arctic also suffered declines, though they were not as extreme as those in the Antarctic.

It was soon found that the cause of the phenomenon was an increase in the atmospheric content of chlorine (and also bromine) which reacts with ozone to produce O_2 + ClO. The rapid increase of chlorine coincided with the industrial production of long lived, chemically inert chlorine compounds, in particular the chlorofluorocarbons $CFCl_3$, CF_2Cl_2 and $C_2F_3Cl_3$ with atmospheric lifetimes of, respectively, 45, 100 and 85 years [51]. These gases were particularly useful because of their inertness in a wide variety of applications: as a pressurized gas in spray bottles, in refrigerators, in fire extinguishers, etc. Because of their long lifetimes in the atmosphere, their abundances have been increasing rapidly. They may be destroyed by sunlight, forming radicals like ClO and later Cl_2O_2 which is broken up by sunlight to produce free chlorine. The ozone destruction is particularly effective on the surfaces of small particles which form as high-altitude polar clouds at temperatures below –85°C [52]. As a result, the depth of the ozone hole is dependent on the stratospheric temperature, which varies from year to year. The record 2006 hole (Figure 5.15) was a consequence of unusually low temperatures, while the much smaller hole in 2002 was the result of a sudden stratospheric warming event. Since the chlorine formation depends on sunlight, the hole forms only when the Sun returns after the polar night; once the solstice approaches, temperatures are too high for stratospheric clouds to form.

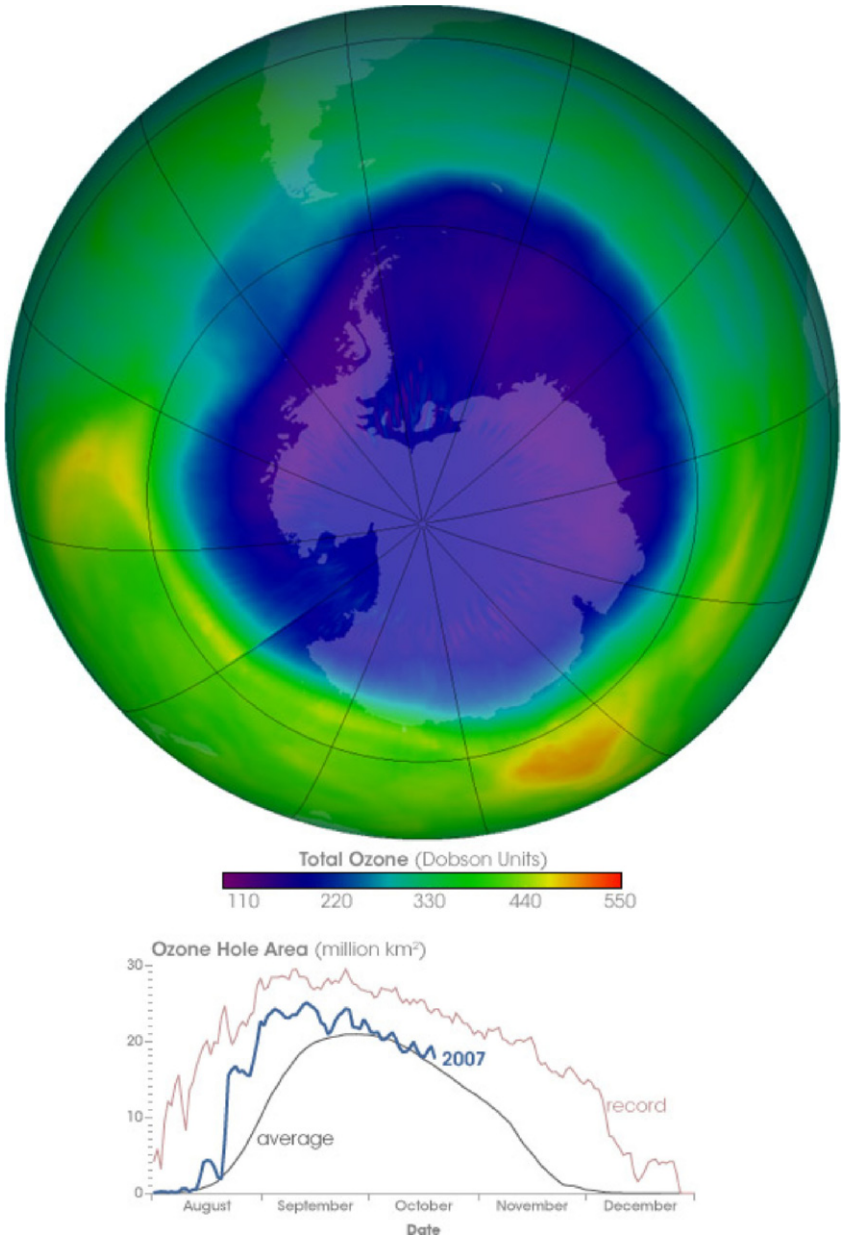


Figure 5.15 The ozone hole over Antarctica in 2007. The figure below the image shows the annual development of the area of the hole during 2007, during the record year 2006 and during more average preceding years. The ozone-destroying reactions require low temperatures and light from the Sun. The hole forms in early Antarctic spring and disappears three months later when the temperatures become too high. (*Source*: Earth Observatory NASA.)

The long lifetimes of the important chlorofluorocarbons have had the effect that their concentrations in the stratosphere are only beginning to diminish long after their production ceases. Moreover, discarded refrigerators, air conditioners and fire extinguishers may retain the gases for a long time before releasing them into the atmosphere. So it is no surprise that the restoration of the ozone layer is a slow process that will only be completed later in the century.

While the stratospheric ozone is beneficial in preventing much of the solar ultraviolet radiation from reaching the surface, tropospheric ozone tends to be harmful to the biological world. Some of it has its origin in the stratosphere, but today much is produced by complex chemical processes from industrial and biomass burning pollution. Since the lifetime of ozone at ground level is very short, it is not well mixed in the atmosphere and high concentrations are frequently found down wind from polluting sources.

The story of the ozone hole is a textbook example of how science-based policy making should function: analysis of possible dangers resulting from human activities, a shocking observation showing the dangers to be real, and the adoption of an international treaty (the Montreal Protocol, Chapter 11) to ensure that disaster is avoided. Perhaps it is also an illustration of the changes of the mood in the world. The troubles of the Kyoto Protocol (Chapter 11) limiting CO_2 emissions show that today such collaborative efforts have become much more difficult to implement.

5.11 Notes and references

[1] Huber, B.T., 1998, 'Tropical paradise at the cretaceous poles', *Science* **282**, 2199-2200.

[2] Tarduno, J.A. et al., 1998, 'Evidence for extreme climatic warmth from late Cretaceous arctic vertebrates', *Science* **282**, 2241–2243.

[3] Lear, C.H. et al., 2000, 'Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite', *Science* **287**, 269–272.

[4] Holmes, A., 1944, *Principles of Physical Geology,* Nelson & Sons, p. 216.

[5] Lea, D.W. et al., 2000, 'Climate impact of late Quaternary Pacific sea surface temperature variations', *Science* **289**, 1719–1724.

[6] Von Deimling, T.S. et al., 2006, 'How cold was the last glacial maximum?', *Geophysical Research Letters* **33**, L14709, 1–5.

[7] Magnusson, M. and Pálsson, H., 1965, *The Vinland Sagas*, Penguin Books, p. 21. The cairn is found at latitude 72°9. In a letter of 1266 there is mention of another voyage that reached 76°.

[8] Magnusson, M. and Pálsson, H., 1965, *The Vinland Sagas*, Penguin Books, p. 23.

[9] Sagarin, R. and Micheli, F., 2001, 'Climate change in nontraditional data sets', *Science* **294**, 811.

[10] Magnuson, J.J. et al., 2000, 'Historical trends in lake and river ice cover in the northern hemisphere', *Science* **289**, 1743–1746.

[11] Oerlemans, J., 2005, 'Extracting a climate signal from 169 glacier records', *Science* **308**, 675–677.
[12] Thompson, L.G., 2002, 'Kilimanjaro ice core records: evidence of Holocene climate change in tropical Africa', *Science* **298**, 589–593; Cullen, N.J. et al., 2006, 'Kilimanjaro glaciers: recent areal extent from satellite data and new interpretation of observed 20th-century retreat rates', *Geophysical Research Letters* **33**, L16502, 1–4. Also in the Rwenzori mountains only 1 km^2 of ice was left in 2003 of some 6 km^2 early in the 20th century as a consequence of rising temperatures, see Taylor, R.G. et al., 2006, 'Recent glacial recession in the Rwenzori Mountains of East Africa due to rising temperatures', *Geophysical Research Letters* **33**, L10402.
[13] See references 19 and 20 in Chapter 8.
[14] Shepherd, A. et al., 2003, 'Larsen ice shelf has progressively thinned', *Science* **302**, 856–858.
[15] Arrhenius, S., 1896, *Philosophical Magazine*, Fifth Series, **41**, 237.
[16] Crowley, T.J. and Berner, R.A., 2001, 'CO_2 and climate change', *Science* **292**, 870-872.
[17] Zachos, J. et al., 2001, 'Trends, rhythms and aberrations in global climate 65 Ma to present', *Science* **292**, 686–693.
[18] Pagani, M. et al., 2005, 'Marked decline in atmospheric carbon dioxyde concentrations during the Paleogene', *Science* **309**, 600–602.
[19] Coxall, H.K. et al., 2005, 'Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean', *Nature* **433**, 53–57.
[20] McPhaden, M.J. et al., 2006, 'ENSO as an integrating concept in Earth science', *Science* **314**, 1740–1745.
[21] Fedorov, A.V. et al., 2006, 'The Pliocene paradox (mechanisms for a permanent El Niño)', *Science* **312**, 1485–1489.
[22] Scher, H.D. and Martin, E.E., 2006, 'Timing and climatic consequences of the opening of Drake passage', *Science* **312**, 428–430.
[23] Jouzel, J. et al., 2007, 'Orbital and millennial Antarctic climate variability over the past 800,000 years', *Science* **317**, 793–796. See also *Science* **310**, 1213–1321, 2005. Petit, J.R. et al., 1999, 'Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica', *Nature* **399**, 429–436. The most recent ice core from Greenland is presented by North Greenland Ice Core Project Members, 2004, 'High resolution record of Northern Hemisphere climate extending into the last interglacial period', *Nature* **431**, 147–151.
[24] Milankovitch, M., 1941, 'Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem', *R. Serbian Academy Special Publication* **132**, vol. 33, 1–633.
[25] Medina-Elizalde, M. and Lea, D.W., 2005, 'The mid-Pleistocene transition in the tropical Pacific', *Science* **310**, 1009–1012.
[26] Zimov, S.A. et al., 2006, 'Permafrost and the global carbon budget', *Science* **312**, 1612–1613. See also Zimov, S.A. et al., 2006, 'Permafrost carbon: stock and decomposability of a globally significant carbon pool', *Geophysical Research Letters* **33**, L20502, 1–5.

[27] Bintanja, R. et al., 2005, 'Modelled atmospheric temperatures and global sea levels over the past million years', *Nature* **437**, 125–128.
[28] Jouzel, J., 1999, 'Calibrating the isotopic paleothermometer', *Science* **286**, 910–911.
[29] Hanebuth, T. et al., 2000, 'Rapid flooding of the Sunda shelf: A late-glacial sea level record', *Science* **288**, 1033–1035.
[30] Cuffey, K.M. and Marshall, S.J., 2000, 'Substantial contribution to sea-level during the last interglacial from the Greenland ice sheet', *Nature* **404**, 591–594.
[31] Kerr, R.A., 1991, 'Global temperature hits record again', *Science* **251**, 274.
[32] Pollack, H.N. and Huang, S., 2000, 'Climate reconstruction from subsurface temperatures', *Annual Review of Earth and Planetary Science* **28**, 339–365.
[33] Jones, P.D. and Mann, M.E., 2004, 'Climate over past millennia', *Review of Geophysics* **42**, 143–185.
[34] Moberg, A. et al., 2005, 'Highly variable Northern Hemisphere temperatures reconstructed from long and high-resolution proxy data', *Nature* **433**, 613–617.
[35] Jones, P.D., 2006, 'Climate over the last centuries from instrumental observations', ISSI Workshop on Solar Variability and Planetary Climates. This is an update of Jones, P.D. et al., 1999, 'Surface air temperature and its changes over the past 150 years', *Review of Geophysics* **37**, 173–199.
[36] Solanki, S.K. et al., 2004, 'Unusual activity on the Sun during recent decades compared to the previous 11000 years', *Nature* **431**, 1084–1087.
[37] Fröhlich, C., 2006, 'Solar irradiance variability since 1978', *Space Science Review* **125**, 53–65.
[38] Foukal P. et al., 2006, 'Variations in solar luminosity and their effect on the Earth's climate', *Nature* 443, 161–166
[39] Bond G. et al., 2001, 'Persistent solar influence on north Atlantic climate during the Holocene', *Science* **294**, 2130–2136.
[40] Stothers, R.B., 1984, 'The great Tambora eruption in 1815 and its aftermath', *Science* **224**, 1191–1197. See also Briffa, K.R. et al., 1998, 'Influence of volcanic eruptions on northern hemisphere summer temperatures over the past 600 years', *Nature* **393**, 450–454.
[41] de Silva, S.L. and Zielinski, G.A., 1998, 'Global influence of the AD 1600 eruption of Huaynaputina, Peru', *Nature* **393**, 455–457.
[42] Newhall C.G., et al., 2002, 'Pinatubo Eruption 'to make grow', *Science* **295**, 1241–1242; Robock, A., 2002, 'Pinatubo eruption, the climatic aftermath', *Science* **295**, 1242–1244.
[43] For 1400–1970 data from MacFarling Meure C. et al., 2006, 'Law Dome CO_2, CH_4 and N_2O ice core records extended to 2000 years BP', *Geophysical Research Letters* **33**, L14810, 1–4. For 1970–2100 data and projections from the IPCC Third Assessment report, *Climate Change* 2001, WGI.
[44] Crowley, T.J., 2000, 'Causes of climate change over the past 1000 years', *Science* **289**, 270–276.
[45] Bengtsson, L. et al., 2006, 'On the natural variability of the pre-industrial

European climate', *Climate Dynamics*, DOI: 10.1007/s00382–006–0168-y, 1–18.
[46] Ruddiman, W.F., 2003, 'The anthropogenic greenhouse era began thousands of years ago', *Climatic Change* **61**, 261–293; see also *Scientific American*, March 2005, 34–41.
[47] Crutzen, P.J., 1971, 'Ozone production rates in an oxygen-hydrogen-nitrogen atmosphere', *Journal of Geophysical Research* **76**, 7311–7327.
[48] Molina, M.J. and Rowland, F.S., 1974, 'Stratospheric sink for chlorofluoromethanes: chlorine atom catalysed destruction of ozone', *Nature* **249**, 810–812.
[49] Farman, J. et al., 1985, 'Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interactions', *Nature* **315**, 207–210.
[50] *Antarctic Ozone Bulletin*, No. 7/2006, World Meteorological Organization.
[51] IPCC (WGI), 2001, p. 244.
[52] Solomon, S. et al., 1986, 'On the depletion of Antarctic ozone', *Nature* **315**, 207–210.

6 Climate Futures

I want to testify today about what I believe is a planetary emergency – a crisis that threatens the survival of our civilization and the habitability of the Earth.

Al Gore

In 1991 there appeared an article [1] by a well-known scientist entitled 'Does climate still matter?' with the summary stating, 'We may be discovering climate as it becomes less important to well being. A range of technologies appears to have lessened the vulnerability of human societies to climate variation.' In 2007 more than a thousand contributors congregated in various places to draft the Intergovernmental Panel on Climate Change Fourth Assessment Report (henceforth IPCC-AR4) under three headings (Table 6.1). As Working Group II concluded [2], 'Some large-scale climate events have the potential to cause very large impacts, especially after the 21st century'. Part of the difference is that the 1991 article was written from a developed country perspective, while it is now clear that climate change will strike the less-developed countries the hardest. But, in addition, we have begun to realize that many aspects of global warming will become irreversible if no action is taken to limit CO_2 emissions during the first half of the present century. Perhaps the clearest example is the risk that an irreversible melting of the ice caps on Greenland and on west Antarctica would be initiated which could raise sea level ultimately by 13 meters, flooding large heavily populated areas. It is doubtful that any foreseeable technological fixes would be able to mitigate such a development.

In the distant past there have been long periods of great warmth and high CO_2 concentrations in the atmosphere. Somewhat ironically, geologists call such a period a 'climatic optimum'; life flourished under the warm, humid conditions. True enough, from time to time species became extinct, but others appeared and, generally, diversity increased. Many species tolerated slow climatic changes rather well, but our present overpopulated world is a different story. There are hardly any areas to which a billion people can flee when conditions, in India for example, become too difficult. Moreover, although the natural world has much capacity for adaptation to slow changes, global warming occurs on a timescale of a century or less. The great speed of the changes makes everything much more difficult. It has become evident, therefore, that all efforts should go towards slowing down the pace of change by dealing with its cause, the increasing rate of

Table 6.1 The Working Groups of the Intergovernmental Panel on Climate Change: Fourth Assessment Report

WG I:	*Climate Change 2007: The Physical Science Basis*
WG II:	*Climate Change 2007: Impacts, Adaptation and Vulnerability*
WG III:	*Climate Change 2007: Mitigation of Climate Change* *Climate Change 2007: Synthesis Report*

These reports are the successors to those in the *Third Assessment Report, Climate Change* 2001 (TAR). Other relevant reports of the IPCC include:

- *Special Report on Emission Scenarios* (2000)
- *Special Report on Carbon Dioxide Capture and Storage* (2005)

These reports have been or will be published by Cambridge University Press. *Summaries for Policy Makers* (SPM) are available on the Internet.

The economic aspects of climate change have been extensively studied in Stern, N., 2006, *Stern Review on the Economics of Climate Change*, Cambridge University Press.

CO_2 production. To evaluate how much of a reduction is needed, we have to understand the relationship between CO_2 emissions and atmospheric CO_2 concentrations and their effects on temperature and rainfall.

6.1 Scenarios for future climates

In the preceding chapter we have seen how models of the Earth's climate have allowed us to understand some of the changes that occurred in the past. The large temperature changes during the ice ages appear to have been related to rather modest variations in the orbit of the Earth around the Sun. In addition, solar variability and volcanic eruptions may have played a modest role, while, more recently, CO_2 and other greenhouse gases due to human agricultural and industrial activities have become of dominant importance. Associated with the increasing temperatures after the last ice age, there has been a more than 100-meter rise in sea level as a result of the melting of the huge ice caps that covered Scandinavia, Canada and other areas. Thereafter further increases have been negligible until recently.

Because the last several millennia seem to have been endowed with a climate that had a rather stable temperature and sea level, we have tended to think that this has been the 'normal' situation. All recorded history of the human race has taken place in this period of stability, even though regional droughts and floods have left anguished memories. But over the past few hundred thousand years that we can study with sufficient time resolution, there are few periods of such stability, and from time to time there are remarkably rapid variations on decadal timescales. So climate is far more dynamic than we had assumed. Now that

human activities are demonstrably changing the atmosphere, it is important to investigate how the future climate may be affected. If we find that the changes have negative effects, it would be reasonable to ask what we can do to diminish these. In the case of the ozone-destroying chemicals, we have already done so. Because the economic interests involved were rather modest, this was achieved with remarkable speed.

In a very general way it is not too difficult to foresee the direction in which the climate will evolve on a timescale of a century. The greenhouse gases, and in particular CO_2, will continue to increase as long as hydrocarbons are used as a source of energy. As a result, the greenhouse effect will be increased and the temperature should rise. The increase in temperature will also affect the oceans and gradually percolate to greater depth. Since water expands when the temperature increases, the sea level should rise. Glaciers and polar ice caps may begin to melt and so further enhance the rise. But to be useful the forecasts have to be more specific: how many degrees of temperature increase and how many meters of sea level rise should we expect, and in how many years will such things happen? Such quantitative questions can only be answered on the basis of climate models (see Box 6.1). With such models we may retroactively 'predict' the climate of the last millennium. Models that do this well stand perhaps a better chance to predict the future, although we cannot exclude that two different models may make the same 'prediction' for the past, but a very different one for the future. After all, the CO_2 concentrations we are beginning to experience are entirely beyond the range of those of the last thousands or even millions of years. Many different models, based on different assumptions on critical processes and parameters, have to be analyzed to see how much of an uncertainty there is.

To study the future evolution of a model we have to specify the external factors that influence it. In the preceding chapter we have seen that the major climate fluctuations of the past million years have been paced by the orbital variations of the Earth with periodicities of tens of thousands of years. So, for the coming few centuries they may be largely neglected, though for the 100,000-year world they will be of great importance. Intrinsic solar variations will continue to play a role on decadal and longer timescales, but are now thought to account for no more than some 4% of greenhouse gas forcing in 2005 [3] (Figure 6.1 – see also reference [4]). For the moment we are unable to predict these substantially into the future, except for those related to the 11-year sunspot cycle. Even less predictable are the volcanic eruptions that generally have effects that persist for only a few years.

What we need first of all is an estimate of the future concentrations of greenhouse gases due to industrial and agricultural practices: we need a time line of the annual inputs of CO_2, methane and other gases, and also of aerosols which, depending on their properties, may keep solar energy out and the Earth's radiation in (or both) and change cloudiness and precipitation. To obtain the input of CO_2, assumptions have to be made about the evolution of the number

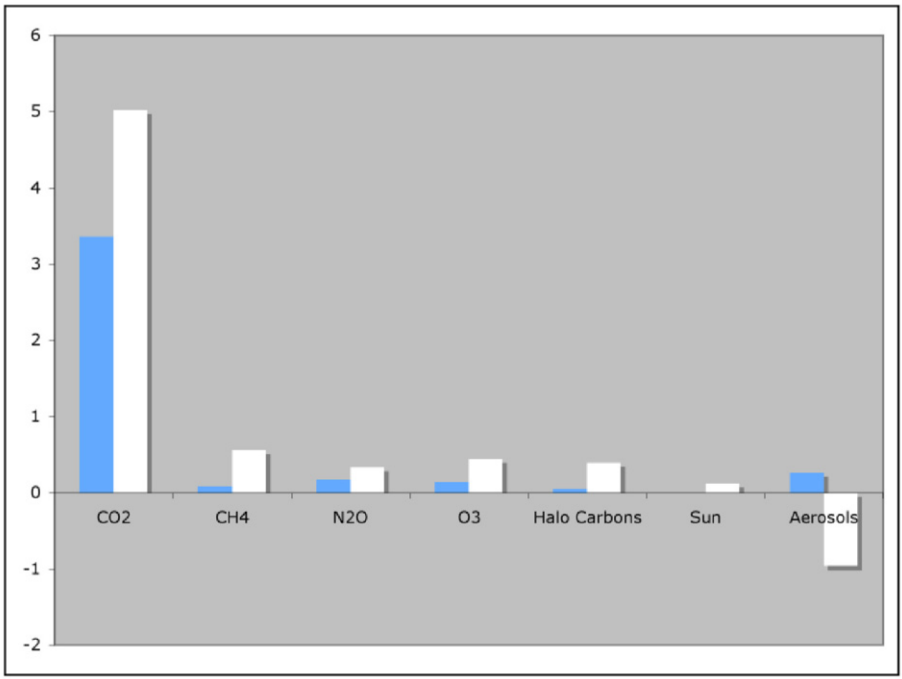


Figure 6.1 Climate forcings. The forcings in watt per square meter from 2005 to 2100 (filled bars) and from 1750 to 2100 (open bars) as predicted from the mean of climate models with the A1B scenario. Only forcings >0.1 W/m^2 are shown. Data for 1750–2005 have been taken from figure SPM-2 in the IPCC: AR4-WGI and for 2005–2100 from Appendix II.3 in the IPCC: TAR-WGI [4]. Aerosol forcings are extremely uncertain. The dominance of CO_2 for the future development of climate is evident. The differences between open and filled bars represent the forcings in 2005 with respect to pre-industrial times, when CO_2 was not yet so pre-eminent and other greenhouse gases, aerosols and a small solar component all played a role.

of people on Earth, on their per capita energy use, and on the sources of energy: hydrocarbons or nuclear and renewables. A further CO_2 contribution – other than methane, the most important one – is of current biological origin: CO_2 is contributed to the atmosphere by deforestation and removed by tree planting, while much of the methane results from the cultivation of rice and the digestive processes of cattle. At present the climate forcing due to CO_2 is about three times more important than that of methane. Currently, other gases and aerosols have effects that largely balance, although there is a lot of uncertainty about aerosols.

The Intergovernmental Panel on Climate Change (IPCC) has developed scenarios for the evolution of industrial and agricultural emissions until the year 2100 [5]. These are grouped into four families: A1, A2, B1 and B2. Different assumptions are made about future population numbers, speed of industrial

Box 6.1 Climate Models

To qualitatively understand past and future climates we need a 'climate model', which gives a mathematical description of conditions in the atmosphere, the oceans, the land surface and their couplings. In the most sophisticated models the temperature, pressure, wind speed and direction, humidity and eventually other items are specified at a large number of grid points over the Earth in the atmosphere. In the most complex models there might be a grid point every degree in longitude and latitude, which corresponds to some 40,000 such points in all. In addition, there would be up to 50 layers at different heights in the atmosphere. The surface features of the Earth would also be specified. In the ocean the grid might also be specified over 50 different layers in depth. So, in total there might be several million points. In the ocean the salinity and the CO_2 content would also be included. While in some models the number of layers would be less and in others the grid spacing wider, it is clear that large computers are required to handle all these data.

Even a $1^\circ \times 1^\circ$ grid corresponds only to typically 100×100 km resolution. So if we want to know if there will be much cloudiness over a certain area, which makes a large difference in the reflectivity to the solar radiation, we need to have a theory of cloud formation and also of rainfall. The surface of the land affects the wind and the wind generates waves on the ocean surface and affects ocean currents. All of these have to be included in an *a priori* parameterized form. Uncertainties in the physics and chemistry of these processes affect the results. But there is more: the organisms that live in the ocean may draw down CO_2 into its deeper reaches and bury the CO_2 as carbonates in the sediments. On land, plants may absorb CO_2 although much of this may be returned to the atmosphere as the plants decay. Therefore, biological processes must also be included in the model.

The length of time over which the model is to be calculated also determines the amount of computing power needed. In practice this means that in many models simplifications have to be made. When we wish to compute such a model over thousands of years, we will employ models with fewer atmospheric or oceanic layers and we may reduce the number of grid points. Thus there is a whole 'hierarchy' of models for different purposes.

The models describe all the feedback processes that occur within the atmosphere–ocean system. But the important 'forcings' of the models – the factors external to the system – have to be specified independently. The external forcings include the changes of the solar energy flux at the Earth and its distribution over the surface, of the aerosols ejected by major volcanic eruptions into the stratosphere, and also of the aerosols continuously introduced into the atmosphere by industrial processes and by dust from deserts, the changes in the concentrations of the greenhouse gases, etc.

Some models have a higher sensitivity to changes in some of the forcings

than others. In discussions of the future evolution of climate a convenient way of characterizing the models is to see how, when all forcings except CO_2 are kept constant, the global mean temperature of the model is changed when the CO_2 concentration is doubled from the pre-industrial value of 275 ppm to 550 ppm. In the IPCC *Third Assessment Report* in 2001 a large number of models is listed in which the temperature increase ranges from 2.0 to 5.1°C. In 2007 [3], with better models, the range had narrowed from 2.0 to 4.5°C with a most likely value of 3°C. While these improved models are welcome, it is also true that modelers will have a natural tendency to introduce similar 'improvements' in the description of processes like cloud formation and cloud properties. Comparisons with the fossil record are therefore important in increasing our confidence in the models.

Table 6.2 Different scenarios for future climate [7].

Subsequent columns give the name of the scenarios, four SRES [5] and three with the CO_2 concentrations, ultimately stabilized at respectively 450, 550 and 1,000 ppmv; the cumulative anthropogenic CO_2 emissions during the 21st century in gigatons of carbon, the year 2100 CO_2 concentrations in the atmosphere in ppmv and the multimodel average of the temperature change between 2000 and 2100. In the SRES the anthropogenic carbon emissions are defined in the scenarios and the resulting CO_2 concentrations calculated from the models, while in the stabilization scenarios the maximum CO_2 concentrations are specified and the corresponding emissions calculated. The results depend somewhat on the adopted time lines for the CO_2 concentrations. We follow here the so-called WRE scenarios which have been optimized based on economic considerations [8]. However, the cumulative CO_2 emissions until 2100 of some alternatives would differ by only 2–3%, a small range compared to that due to uncertainties in the climate models. Note that 550 ppmv of CO_2 corresponds to double the pre-industrial value and that in 2006 the CO_2 concentration had reached 380 ppmv. The ΔT's would increase a further 0.5°C after 2100 even if the emissions were stabilized at 2100 values.

As the climate warms the feedbacks in the carbon cycle would tend to further increase the CO_2 concentrations and thus to reduce the emissions allowed if a certain concentration limit is to be respected. In the second column are given in parentheses the corresponding reduced values of the emissions in some illustrative cases from the IPCC: AR4-WGI. These feedbacks remain very uncertain.

Scenario	$\sum$ CO_2(gigatons C)	CO_2 (ppmv)	ΔT (°C)
A1B	1360	700	2.7
A2	1700	840	3.5
B1	910	540	1.7
B2	1090	610	2.4
WRE 450	670 (490)	450	1.8
WRE 550	960	540	2.3
WRE 1,000	1,420 (1,100)	*Stabilization only in 2375*	

development, etc. While a total of 40 different scenarios have been constructed, four 'marker scenarios' have been mainly utilized in the climatological literature: A1B, A2, B1, B2, with A1B the middle case of three A1 scenarios with a certain balance between the use of fossil fuels and renewable energy sources.

The A1B and B1 scenarios have a world population that peaks in 2050 at 8.7 billion people and declines to 7 billion 50 years later. This seems to us to be a remarkably optimistic assumption. A1B is a scenario with rapid economical growth, high consumption, widespread education and technological progress. In the B1 scenario there is more emphasis on environmental and social matters and also a strong commitment to education. A2 has a relatively rapid population increase to 15 billion people in 2100, slower technological, social and environmental improvements, and much global inequality. Finally, the B2 scenario follows the UN medium variant for population growth to 10.4 billion in 2100: there is much concern for education and the environment and less so for technological innovation.

From such scenarios the IPCC *Special Report on Emission Scenarios* (SRES) has been constructed [5]. According to the IPCC–SRES, all scenarios 'are equally sound' [6]. Nevertheless, scenario A2 appears, to us, to be particularly unattractive and B1 is perhaps optimal. However, the most important aspect of the scenarios is the total greenhouse gas production during the century with the precise time line having a much smaller climatological effect than the uncertainties in the climate models. Once the time line has been adopted, the climate model determines what the temperature increase and the CO_2 concentration will be. Independently of the scenario chosen, the temperatures increase by 0.2°C per decade until 2020. By 2050 the average is 1.3°C above 2000, with A1B 0.2°C higher and B1 0.2°C lower. Thereafter the differences increase (Table 6.2) [7]. It should be noted that after 2100 the temperatures would continue to increase, even if CO_2 production stopped, because of the inertia of the oceans. This would yield a further increase of at least 0.5°C.

The SRES scenarios cover a wide range of possible futures depending upon the socio-political assumptions that are made about CO_2 emissions (see Section 6.7). The scenarios themselves are not particularly important, it is only the time line of the CO_2 emission. Alternative scenarios have been developed in which the final CO_2 concentrations are specified and the corresponding emissions calculated. Since, in principle, the CO_2 concentration determines the temperature, the final temperature increase that is considered acceptable or unavoidable may be specified. The only uncertainties are due to the climatological models, and, to a lesser extent, to the initial time line of CO_2 concentrations. Here we follow the WRE time lines [8].

The relationship between the amounts of CO_2 emitted, the CO_2 concentration in the atmosphere and the temperature increase are model dependent, and so the precise values are still quite uncertain. For example, the CO_2 concentration for the A1B scenario ranges from about 600 to 900 ppmv depending upon the climate model chosen. The temperature increase would range between 1.9°C and

3.5°C for seven different models. So while the temperature increase is undoubtedly larger in scenario A1B than in B1, the average modeled values for ΔT could still be rather far off.

6.2 Geographic distribution of warming

The warming to-date has been very non-uniform over the Earth and the same is expected for the future. Two general features stand out: warming over land is stronger than over the oceans, and over northern regions it is much stronger than nearer to the equator. These features are shared by essentially all models of climate and seem to be largely independent of the global average warming that they predict. In fact, a multi-model average of the 20 models considered in the 4th IPCC Assessment Report yields a mean ratio of land to sea warming of 1.55 with a range of 1.36 to 1.84 for the most divergent models [9]. Taking into account the areas occupied by land and by oceans, this corresponds to the land warming 34% more than the global average and the oceans 14% less. Since people live on the land, the effects of global warming are therefore still more important than one would have thought from the global figures in Table 6.2 alone. It is believed that the effect is due mainly to the fact that the evaporation over the oceans depends rather strongly on temperature. The heat energy needed for the evaporation reduces that available for heating the air, an effect that is largely absent over land [9]. The large heat capacity of the oceans may also play some role until a new equilibrium at larger CO_2 concentrations has been established.

The amplification of the warming over northern areas is expected to be even larger [10]. For the lands north of 45–50° latitude the models predict on average a warming some 90% above the global average during winter and 40% during summer. At first it was thought that the northern warming was enhanced by the 'snow albedo effect': when warming makes the snow melt, the darker underlying soil is exposed and more of the Sun's light is absorbed instead of being reflected back into space. However, while this is a factor, the reality is much more complex and not yet fully understood [10]. In fact, also during warm periods in the geological past, when little snow could have been expected, the Arctic warming was particularly strong.

Much of the rest of the world's land is expected to have a warming within 10% of the global *land* average, with the tropics generally being on the low side and the intermediate latitudes on the high side of that range [11]. South America below Amazonia, South Australia and the islands of South-East Asia would have warming below the global land average. The liveability of an area may be particularly affected by the highest summer temperatures. The Mediterranean region, central Asia, the Sahara region and west and central North America are expected to have summer temperatures 10% or somewhat more above the global land average, i.e. some 50% above the global average [11]. In the case of the Sahara region, and based on the A1B emission scenario, summer temperatures

would then average 33–34°C. The reliability of regional forecasts is perhaps open to some doubt, since the differences between the models are still large.

Perhaps even more important than the temperatures are the expected precipitations in the different regions. Droughts in dry areas can have particularly catastrophic effects as has been amply demonstrated by the perishing of numerous civilizations when the water gave out (see Section 4.6.3). Since warmer air can contain more water vapor, the overall effect of global warming should be an increased global precipitation. The global effect is not very large, some 4% or so. In the northern areas the strong warming is expected to be accompanied by precipitation increases of the order of typically 15% by 2100 on the A1B scenario [11]. But in the subtropics, where dry air descends from higher up, a further drying is predicted by most climate models. Particularly hard hit would be the Mediterranean area, the Sahara region and Central America (essentially Mexico) where reductions of 10–15% in annual rainfall are predicted in 2100. More modest reductions are envisaged in central Asia and in the southernmost parts of the continents in the southern hemisphere. However, these reductions of around 5% result from averaging 19 models, some of which also predict the opposite effect. In the case of the Mediterranean area and Central America the greatest reductions are of the order of 20% in the dry season with a remarkable unanimity of virtually all models that there is a reduction [11]. The results are sensitive to the definition of the regions considered. For smaller subregions the effects may be stronger. For example, the same 19 models indicate for the southwestern USA (Texas and California) an annual reduction of 50 mm in rainfall [12]. This is now a dry semi-desert region which, at present, lives in part on a finite supply of fossil underground water that will probably run out before the end of the century. In fact, just as the people of more northern climes have begun to move south and discovered the pleasures of air-conditioned life in the subtropical deserts, there is a risk that the water may run out there (Section 8.1).

In addition to the mean changes of temperature and precipitation, the interannual variability of these quantities is of great importance. In most cases it appears that local interannual climate variability increases as the Earth warms, and heat waves, floods and droughts are likely to become more frequent.

Generally, if we consider the climate system on, say, the first of January, we find that at the end of the year it is somewhat different owing to random fluctuations that have accumulated during the year. As a result, the next year begins with an altered state, so the evolution of the system is also different. Depending on the degree of randomness, the average over the year may also be rather different. Thus one hot year may be followed by another, or it may be cold. If the climate system is stable, it will return after some years to the same mean situation as before. If we wish to predict future conditions with a climate model, we should integrate it over a few decades to find out, by averaging, what the evolution of the main state will be. At the same time we can determine the fluctuations around the mean. When the fluctuations of temperature and

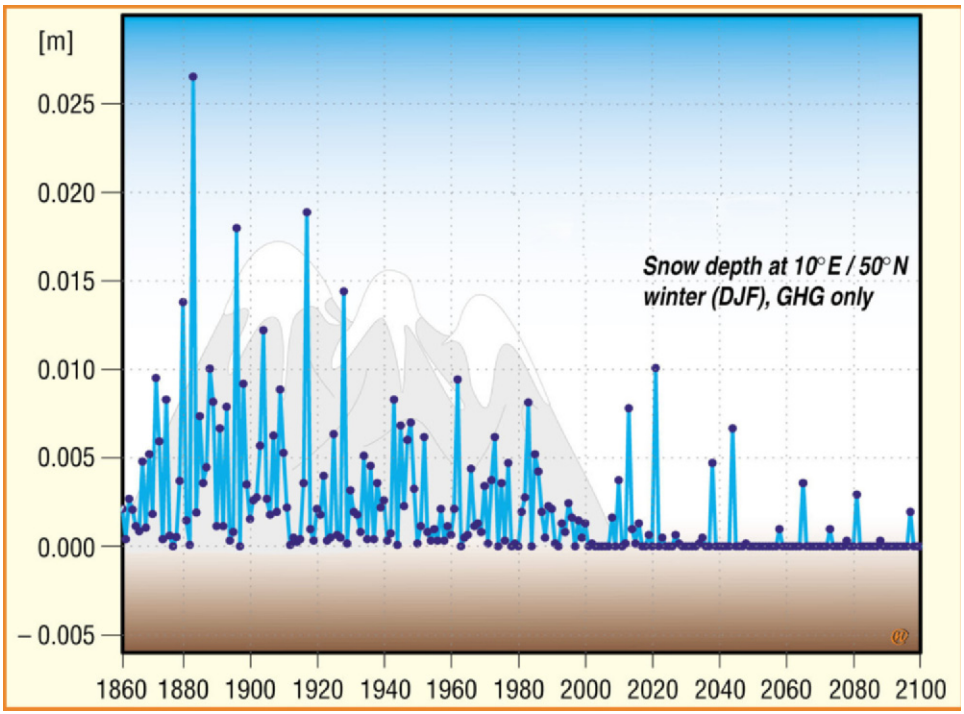


Figure 6.2 Snowfall in one run of a climate model. The interannual variations in the model sometimes dominate over the gradual change induced by increasing concentrations of greenhouse gases. (*Source*: L. Bengtsson.)

precipitation are large, they can have significant effects on the biological world. If important droughts occur from time to time, the vegetation will be very different from that of a non-varying climate with the same average rainfall. After all, what is killed during a drought need not return when next there is flood. So people living in highly variable climates are more exposed to episodic food shortages; one cannot compensate for the hunger of one year by eating twice as much during the next.

The regions most exposed to catastrophic droughts as the climate warms will be those where rainfall diminishes and, at the same time, the variability increases. The larger Mediterranean region will be particularly at risk. For an average A1B scenario, precipitation is expected to diminish by the order of 15% towards the year 2100 and its interannual variability increase by 30%. Hence, in a bad year very little rain would remain. Mexico would be hard hit, while central Asia, southern Australia and South Africa would also become more drought-prone. Important multi-decadal regional droughts appear to have greatly affected the Maya empires in Yucatan, the Pueblo Indians in southwestern USA, the Indus civilization in what is now Pakistan and others, like the dust bowl of the 1930s in the American midwest. But even longer droughts are known: the

Sahara was green and full of animal life some 7,000 years ago and in only a few centuries became what it is now. According to some models this was the result of a random variation going out of hand and leaving it in another equilibrium state [13]. Will such regional events become more likely in a warmer world? We really do not know.

The large variability in the climate also makes it difficult to identify exceptional years as due to global warming. In 2003 a heat wave struck western Europe with, according to some statistics, up to 70,000 fatalities. The temperature for June–August was at least 2°C above the average for the 20th century [14]. Was this due to global warming or was it just an exceptional event of a kind that happens once every few centuries, or both? Another illustration is seen in Figure 6.2 where the snowfall in some area is calculated year by year for a climate model that is forced with slowly increasing greenhouse gas concentrations. One sees that in the model run, snow rich years are gradually diminishing. Nevertheless, from time to time there is a year in which snowfall is higher than it was some decades before, and some people might conclude that warming has stopped. Of course, this is a regional effect. The average global climate is much less variable and the conclusion that it is warming is robust.

6.3 Sea level

Global warming causes the sea level to rise because water expands as it warms. Since it takes several centuries for the surface warmth to penetrate into the deeper reaches of the oceans, this is a rather slow but persistent process. In addition, the warming may cause glaciers, Arctic ice caps and the large ice sheets on Greenland and Antarctica to melt unless snowfall also increases. The sea level appears to have been more or less unchanged since Roman times, beginning to rise slowly during the past century [15]. The current rate of sea level rise has been about 3 millimeters per year during the 1993–2003 period, which is about twice the rate of the 30 preceding years [16]. While the uncertainties in these figures remain non-negligible, this suggests that these processes are accelerating as the Earth warms. Current estimates suggest that half of the 10-year value is due to the warming of the oceans, the other half coming from melting ice. The contribution from the Greenland Ice Sheet has been estimated as no more than 0.2 mm/year, with that of Antarctica perhaps comparable, but still very uncertain. Such ice sheets gain ice mass by snowfall in the higher central areas. From there the ice flows down slowly towards the edges. Along the way it may melt, or in coastal areas lead to the calving of icebergs, which may still travel far from their places of origin.

The Greenland Ice Sheet (GIS) is up to 3–4 km thick; if melted completely it would raise the sea level by about 7 meters. Until recently the ice sheet on Greenland was almost in equilibrium, with the snowfall on the top balanced by the losses at the edges. When the temperature increases a little, a new balance may be established. However, models show that if the local summer temperature

increases by around 2.7°C, this is no longer possible. Unfortunately, climate models suggest that the warming at Greenland will be well above the global average. Even in a scenario with CO_2 stabilization at 550 ppmv, the summer temperature could ultimately increase by 3.8°C and the ice cap would begin to melt away [17]. Because of the immense quantity of ice, complete melting will take time; if CO_2 is stabilized at 550 ppmv, one-third of the ice would have melted by the year 5000, and at 1,000 ppmv essentially all of it with a consequent sea level rise of 7 meters. Of course, it should be remembered that there is still much to be learned about the dynamics of ice sheets. In fact, recent studies seem to indicate that ice sheet behavior is far more dynamic than previously thought [18]. Therefore, the sea level might rise much faster than predicted.

There is evidence that the ice loss in large parts of Greenland is accelerating. Altimeter results show a mean loss of 60–110 km^3/year for the five-year period until 2004, at least double the loss for the preceding five years [19]. Two glaciers speeded up by factors of 2–3 and over their whole catchment area lost a total of 120 km^3/year over five years until 2006, corresponding to a sea level rise of 0.3 mm/year, 50% more than the rise from all of Greenland in the period 1993–2003 [20]. Glacial earthquakes, which have been detected from coastal areas in Greenland, are caused by sudden movements of the ice [21], and the number of such events more than doubled from 2001 to 2005. Thus there seems to be much evidence for a speeding up of the ice loss, perhaps due to surface melt water reaching the glacier bottom through cracks in the ice and lubricating the ice flow. Seeming confirmation of an even larger acceleration of ice loss came from the GRACE satellites which, in principle, measure the gravity variations due to the ice sheet and so its mass changes rather directly (see 'Gravimetry satellites' on page 331). According to these data, from spring 2002 to spring 2004 the ice loss was 100 km^3/year, while during the following two-year period it had reached 340 km^3/year [22]. Distributed over the world's oceans a loss of 340 km^3/year would correspond to a rise in sea level of 0.8 mm/year. However, a revaluation of the errors has shown that these are larger than previously expected [23]; with the errors now ± 150 km^3/year the reality of these variations remains in doubt, although the method remains promising for the future.

As all the evidence for a rapidly increasing ice loss from Greenland pertains to the last decade, it remains somewhat uncertain which part of it could be due to natural variability. However, there are danger signs in the geological record which indicate that in a warmer climate significant amounts of ice could melt. During the last interglacial, the temperatures in the Arctic appear to have been 3–5°C warmer than during recent times [24], and the sea level was probably 4–6 meters higher. This would indicate that part of the GIS melted, though not all of it.

The last interglacial began about 130,000 years ago due to a favorable orbital situation which resulted during April to June in solar forcing of 40–80 W/m^2 at the North Pole even although the mean annual forcing was no

more than 5 W/m^2 [25]. The warming on Greenland is quite sensitive to conditions during early summer. The resulting snow melt then amplifies the warming and melting during the whole summer. These conditions lasted for several thousand years and caused the Greenland ice cap to become substantially smaller and sea level to rise by some 2.2–3.4 meters. The height of the ice sheet was probably reduced by no more than some 500 meters as evidenced by the isotope ratios at the summit. The configuration was then that of a smaller ice cap with steep edges. Since the observed sea level rise was some 4–6 meters, probably an additional contribution came from Antarctica.

The speed of the sea level rise during the last interglacial is still uncertain. Because of the even larger orbital forcings at the time, values in excess of those at the termination of the last glacial period seem possible; these correspond to 11 mm/year. There is some controversial evidence that values of 20 mm/year were attained at some times during the last interglacial, i.e. 1 meter in 50 years. The present warming is expected to be comparable to that during the last interglacial and perhaps comparable rapid rates of sea level rise cannot be entirely excluded [24].

The effect of an open Arctic Ocean on the GIS and the northern climate in general is not entirely evident. Currently, the Arctic is a very dry place with rather low snowfall. Would it increase if the sea ice is gone? During the cold winter most of the Arctic Ocean freezes over, while during the summer the long days melt part of the ice. Around 1980 when satellite data became available, the ice extent in late winter was around 16 million km^2 and at the end of the summer about half as much. Twenty-five years later, by 2005, the winter had diminished by 10%, but the summer ice was reduced by about 25%, while the remaining ice had also thinned [26, 27]. Even more spectacular, two years later summer ice had diminished by a further 20% to about 4.2 million km^2. As a result, the fabled Northwest Passage through the Canadian Arctic had become navigable for the first time in recorded history (Figure 6.3). So the prospect of an ice-free Arctic Ocean seems plausible.

The Antarctic Ice Sheet (AIS) consists of a western part (WAIS) and an eastern part (EAIS). If fully melted the WAIS would add 6 meters to the sea level and the EAIS some 60 meters. The East Antarctic Ice Sheet appears to have been in existence for millions of years. Its first origin 34 million years ago was related to the global decline of CO_2 concentrations over the last 50–100 million years (Section 5.3). A contributing factor may have been the opening up of the channels between the southern continents and Antarctica by a continental drift with the consequent reinforcement of the circumpolar currents, which isolated it from the climate system of the rest of the world. It has therefore had ample stability. Models suggest that warming of more than 20°C would be required to initiate significant melting.

The WAIS is likely to be much less stable because it rests on solid rock mainly below sea level [28], and warming oceans could directly erode the ice. In various places the Antarctic ice extends over the ocean, forming floating ice shelves that are thought to buttress the glaciers further inland. Evidence of potential

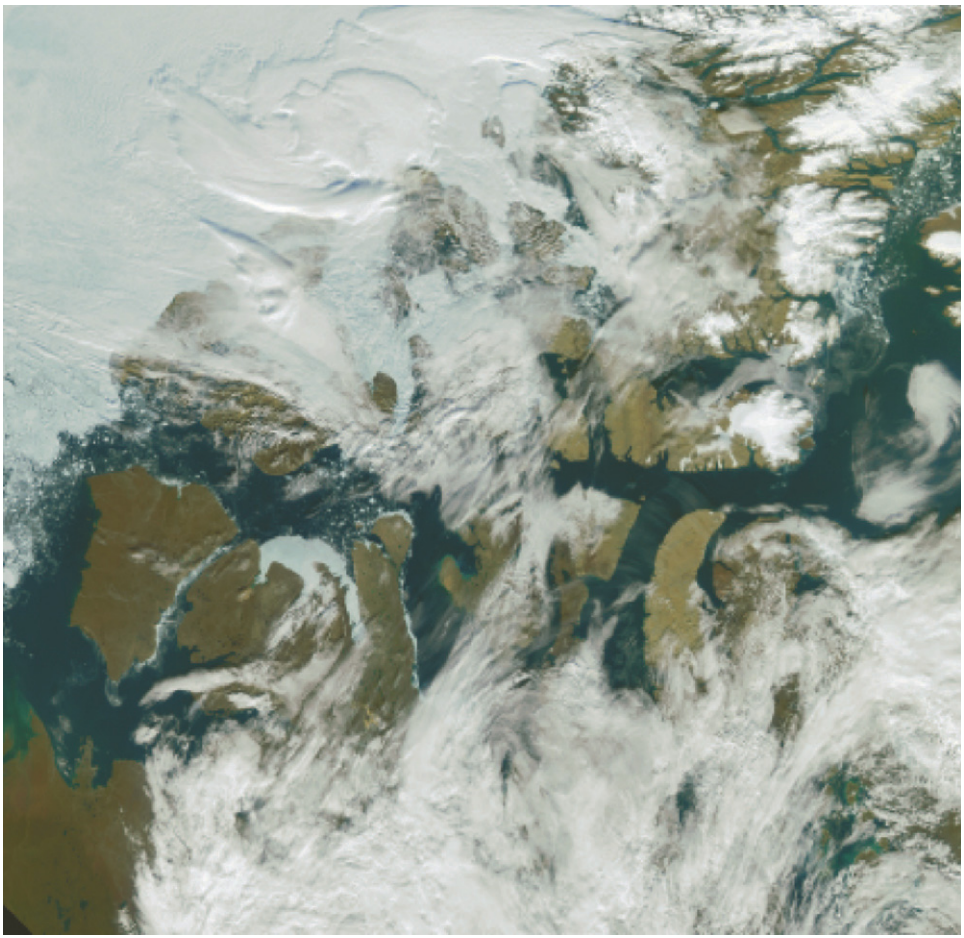


Figure 6.3 The spectacular reduction of Arctic sea ice. As a result, a ship was able to pass from eastern to western Canada. (*Source*: NASA.)

instability has come from the ice shelves around the Antarctic Peninsula, which have retreated by some 300 km^2/year since 1980. In 1995, and again in 2002, large parts of the Larsen ice shelf, 2,000 respectively 3,200 km^2, disintegrated in less than a month. Observations with the European Remote Sensing satellite radar altimeter suggest that this resulted from a progressive thinning of the ice by up to 2–3 meters per decade, perhaps as a consequence of rapid warming in the area of some 2.5°C over the last 50 years [29].

Again, the geological record contains some danger signs. Diatoms – microscopic algae with siliceous cell walls – in a drill core under the WAIS show that at some time during the last million years there was open ocean since these organisms cannot live under the ice [30]. So at least some of the WAIS

must have disintegrated at a time that CO_2 concentrations in the atmosphere were below those of today. Unfortunately, more accurate dating is not yet available. Also, it has been shown that some of the ice shelves that are now crumbling had been in place for at least 10,000 years, indicating that the present events are exceptional [31]. The high sea level during the last interglacial suggests that, in addition to Greenland, another source of melt water was present which probably was the WAIS. For the moment it looks like both the GIS and the WAIS partially survived the interglacial warmth. The rate of melting of the WAIS is still unknown.

6.4 The 100,000-year climate future

The last interglacial period occurred about 130,000 years before the present, which is very comparable to our 100,000-year period. It had lasted not much longer than the 10,000-year duration of the current interglacial, the Holocene, when the temperature fell below present-day values and the slow highly variable decline into full glacial conditions had begun. Some scientists then predicted that the Holocene should also soon be coming to an end. In fact, following the medieval warm period, a slow temperature decline had begun which had led the world into the Little Ice Age. Subsequently in the first half of the twentieth century climate warmed, but by 1950 this had stopped and fears of another ice age surfaced again. However, some two decades later CO_2 and methane concentrations were rapidly increasing and a steep warming had started.

Even if there had been no anthropogenic greenhouse gases, the last interglacial (the Eemian period) would not have been the best comparison for the Holocene, since the Earth's orbital configuration was rather different. In fact, the eccentricity of the Earth's orbit is becoming very small and will remain so for the next 50,000 years or so [30]. As a result, the Milankovitch-type forcing will be much weaker and the evolution of the climate rather different. To find a comparable orbital situation, we have to go back four glacial periods to about 400,000 years ago. That interglacial lasted much longer than the three that followed, as may be seen from the temperatures derived from the deuterium isotope record at Antarctica (see Figure 5.5). After a rapid warming from a deep glacial minimum, temperatures very similar to those today prevailed for more than 20,000 years.

Our knowledge of the Earth's orbit and of the inclination of its axis is sufficiently firm to predict the insolation for any point on Earth for any day for more than 1,000,000 years into the future. In our discussion of the ice ages we have seen that the waxing and waning of the northern ice sheets was directly connected to the summer insolation at high northern latitudes. Figure 6.4 shows how this insolation varied in the recent past and how it will evolve in the future. The exceptionally high insolation 130,000 years ago led to the melting of huge amounts of ice, including perhaps half of the Greenland Ice Sheet and also some of the West Antarctic Ice Sheet. Sea level was some 5 meters or more above

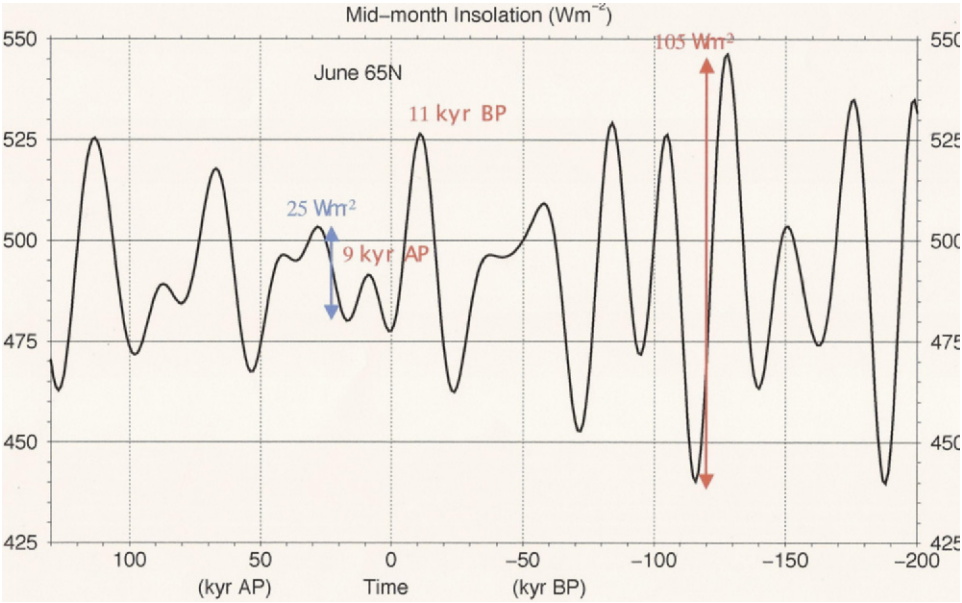


Figure 6.4 Past and future insolation during early summer at 65°N. The exceptionally high insolation some 130,000 years before present (BP to the right) caused a very rapid exit from the previous ice age, but the interglacial was soon terminated by the following deep minimum at 115,000 years BP. During the coming 50,000 years (after present, AP) only rather small insolation variations should occur because the Earth's orbit will be nearly circular. (*Source*: M.F. Loutre.)

contemporary values and temperatures were higher by several degrees (see Chapter 5). But soon thereafter a very deep insolation minimum followed which initiated the next ice age. Subsequent variations were insufficient to remedy the situation until the insolation maximum 11,000 years ago restored interglacial conditions. The subsequent decline may have been almost sufficient to end the Holocene but was not quite sufficiently deep to reach glacial conditions. Also during the next 100,000 years the Milankovitch effects will continue to be important. The future evolution over the coming 50,000 years shows that northern insolation will remain above current values and so no glacial period would be expected to begin. Only some 55,000 years into the future will a somewhat deeper insolation minimum occur that could have the potential of initiating the next ice age. More detailed calculations appear to confirm this [32]. This could have happened if the atmosphere had remained in its natural state with CO_2 concentrations of 280 ppmv during interglacials, but with current CO_2 concentrations at 380 ppmv – much larger than in past interglacials and still rising – the outcome should be very different.

The CO_2 emitted by anthropogenic activities enters into the atmosphere, but at the moment only half of it remains there. The other half is stored in the oceans and in the biosphere on land. The CO_2 in the surface layer of the ocean

rapidly establishes an equilibrium with that in the atmosphere. It takes several centuries for this to percolate into the deeper layers, but once this has happened the CO_2 reacts with carbonates in the sediments at the bottom of the oceans. This process, however, may take some 10,000 years during which atmospheric CO_2 excess slowly decreases to values of the general order of 10% of the initial value, and it may take 100,000 years before the natural carbon cycle takes this up. In fact, the events during the PETM (see Box 6.2) correspond well to such a course of events [33, 34]. It follows that posterity will experience the effects of our CO_2 emissions a long time into the future.

Box 6.2 The Paleocene–Eocene Thermal Maximum (PETM)

A singular climatic event occurred that was associated with an extinction at the boundary of the Paleocene and Eocene epochs 55 million years ago. Suddenly, in less than 10,000 years, tropical sea surface temperatures shot up by some 5°C and at higher latitudes by nearly double that [33]. At the same time the isotope ratio $^{13}C/^{12}C$ decreased significantly. The most plausible interpretation of this event is that biogenic methane hydrates (which have low ^{13}C) in the ocean destablized. These hydrates consist of crystals of water-ice and methane and are stable at low temperatures and high pressures. A warming event, perhaps associated with volcanism, would have freed the methane which was oxidized to CO_2. To obtain the observed low $^{13}C/^{12}C$, some 1,000 to 2,000 gigatons of carbon would have been required. An analysis of carbonates at different depths suggests that three times as much CO_2 was liberated, which would require a supplementary source of CO_2 with higher ^{13}C content [34]. The whole event lasted more than 100,000 years, after which the preceding temperatures and $^{13}C/^{12}C$ ratios were restored [33]. It has been suggested that during the first part of the PETM there were several volcanic events, in which case the recovery time was perhaps no more than 50,000 years.

It is interesting to compare this event with the present anthropogenic perturbation of the atmosphere. It may be estimated that some 500 gigatons of anthropogenic carbon had been produced by 1990. Adding to this the amounts to be emitted by 2100, it is found that in scenario A2 (see Table 6.2) the total would become 2,200 gigatons and in B1 1,410 gigatons. So the 'anthropogenic event' is qualitatively comparable to the PETM. While the PETM occurred in a rather different constellation of the Earth system with warmer temperatures and less ice, if any, perhaps the most interesting aspect is that it took some 50,000–100,000 years for CO_2 concentrations and temperatures to fully return to anterior values. A similar time may be required for the anthropogenic effects to disappear after the ocean and atmosphere have come to a new equilibrium.

Actually the reality is still much more complex. When CO_2 concentrations increase, the oceans become more acidic and, of course, global warming also pervades the deeper reaches. Both factors reduce the capacity of the ocean to take up CO_2. Also, feedbacks on land may have large effects – in particular the melting of the permafrost. Estimates of the carbon that this could release into the atmosphere go up to 1,000 gigatons, which could be comparable to the anthropogenic carbon from the emissions in Table 6.2. We do not know how long the melting of the permafrost would take, so if we could cap the CO_2 concentrations at 450 ppmv and cease producing new CO_2 thereafter, much of it might well survive. After many thousands of years much of the anthropogenic CO_2 would have been taken up by the oceans, and with only 10% of the maximum remaining, the excess would be only 17 ppmv above the pre-industrial 275 ppmv. In that case some 50,000 years into the future the insolation minimum could still suffice to produce the next glacial period. But if we do not manage to stabilize the CO_2 concentrations at such low levels, the melting of the permafrost could lead to much higher concentrations. The higher temperatures would also melt the ice on Greenland, in the WAIS and also most of the sea ice. This would further reduce the reflection of solar light into space and so contribute to additional warming. With CO_2 concentrations reaching values above 1,000 ppmv, the long-term concentrations would undoubtedly suffice to avoid the coming of another ice age.

There would therefore be two possibilities. In the first scenario, which requires the end of anthropogenic CO_2 production quite soon, the ice sheets would remain and present-day conditions might last for several times 10,000 years. Halfway through the 100,000 years glacial conditions would have been re-established [32]. Another scenario would be one with more or longer CO_2 production and with the disappearance of the Greenland and West Antarctic Ice Sheets and possibly a more permanent switch to a 'greenhouse climate'. It is tempting to think that it would be possible to avoid the melting of the ice caps, but to pump later just enough CO_2 into the atmosphere to avoid another glacial period. In any case long-term monitoring of the atmosphere is of essential importance in order to be able to intervene, if necessary. But if one sees the difficulties encountered in something as simple as the Kyoto treaty (see Chapter 11), then achieving general agreement on active intervention will be a very difficult problem. Kyoto imposed modest constraints to achieve aims that almost everyone could agree upon and would not lead to an obvious climate deterioration anywhere. Worldwide climate engineering is a different matter (see Chapter 9). If to avoid another ice age, the Greenland ice cap would have to be melted, one could hardly expect the inhabitants of low-lying countries such as Bangladesh or the Maldives to agree. Moreover, at least for the moment our limited knowledge of atmospheric and oceanic dynamics would make it difficult to reassure the world that no unforeseen problems or disasters would occur. At the same time this is not a reason not to take measures that will improve our collective destiny. But before doing so, we should ensure that we have a full understanding of what we are doing.

6.5 Doubts

While simple climate models have been doing a respectable job in reproducing the temperature record of the last one or two thousand years, it should be noted that this was a period of slow climate changes with an amplitude in temperature of probably less than 1°C from the mean. Moreover the past temperatures still have much uncertainty (see Figure 5.11) and so do the amplitudes of the solar and volcanic influences. Perhaps the fact that the models can be made to represent the past should not be overstressed. If we go a bit further back to the Younger Dryas, some 12,000 years ago, very rapid cooling and warming have occurred with amplitudes of more than 10 degrees and sometimes on timescales of no more than decades or even just years. While this particular event may have had a connection to the huge amounts of melt water associated with the end of the ice age, it does show that in the climate system there may be certain thresholds where a sudden non-linear change occurs that may not be reversible without a large change in the forcing factors. A simple example is the Greenland Ice Sheet: at present it is being fed by snow at 3,000 meters altitude, but when it has gone, the precipitation would fall about at sea level in much of the area. Because of the lower altitude, more snow would have become rain and so a *much* colder climate would be needed to restore the ice cap.

Unfortunately we do not know what other thresholds there are in the climate system. When the temperature augments still a bit further, will we become locked into a permanent warm greenhouse climate such as the Earth has experienced during much of its past? Furthermore, temperature is not the only parameter, and perhaps not even the most important one. Many civilizations have perished not by being a degree warmer, but by persistent droughts. The present-day Sahara dates only from 7,000 before present, before which time more humid conditions prevailed [13]. However, global temperatures had not changed much at that time. Is there a prospect of other areas drying out catastrophically in the future warmer climate?

Melting of the ice sheets is a slow process, but could whole stretches of ice slide down into the oceans and so speed up the process and qualitatively accelerate a rise in the sea level? We know that during the last interglacial the sea level was 4–6 meters, or more, higher than today, but we do not know how long it took to reach that stage. Of course, all such issues may look like climatological curiosities. But with food security uncertain, the loss of large stretches of agricultural land in an overpopulated world could have grave consequences. We shall return to these issues in Section 8.2.

Even in the models considered by the IPCC there remains much uncertainty, which is reflected in the range of predictions for climate change by the year 2100. Of course we shall only know at that time what the quality of the different models really was. It is therefore instructive to look now at how well the present conditions were predicted some 16 years ago. It turns out that the temperature increase is within the range of the predictions, but close to the upper end, while the sea level follows the very upper limit of the predictions [35]. While it cannot

be excluded that the intrinsic variability of the climate system plays a role here, it suggests that the predictions based on the average of many models may underestimate the future increases.

6.6 Consequences of climate change

Global warming will affect almost every aspect of life on our planet. Adopting for the moment the A1B scenario until the year 2100, and no further CO_2 emissions thereafter, by 2200 global temperature would increase by 3.2°C for the multi-model average. Average warming over land would be a third higher and would amount then to 4.3°C, a bit less in the tropics and much more in the Arctic (8°C?).

In a recent interview with America's National Public Radio, Michael Griffin, the Head of NASA, when asked about global warming said,

> *'I guess I would ask which human beings – where and when – are to be accorded the privilege of deciding that this particular climate we have right here today, right now is the best climate for all other human beings. I think that is a rather arrogant position for people to take.'*

Such remarks show a fundamental misunderstanding in high places of the climate issue. It could very well be that another climate could have positive aspects. But since humanity has constructed its complex society under present climate conditions, and since cereals and other agricultural products have been developed under such conditions, a changing climate poses many problems. However, what would make the present crisis particularly acute is the unprecedented speed of climate change. We have seen that very different climates have prevailed over the millions of years of the geological past and that the sea level has varied over many meters. It could be argued that a green Arctic could add as much land as would be lost from the corresponding rise of sea level. But Amsterdam and London, and numerous other cities, have been built for the present level and the costs of moving them to higher sites would be exhorbitant. So, in a sense, the evolution of human society over the last few centuries and millennia has locked us into a situation in which the present climate is actually the optimal one. This does not mean that we cannot adapt ourselves to an unavoidable change, but the faster the change, the more difficult would be the adaptation and, of course, there are limits. Had the temperature been 5°C warmer 50,000 years ago, early humans would have settled in Siberia rather than in India. But, as we noted before, this does not mean that now, if such a temperature increase were to occur, we could move a population of a billion to Siberia.

Several places in the subtropics attain temperatures that approach the limits that humans can tolerate during at least some weeks each year, and evidently a 4°C supplement will push this over the limit. Air conditioning might solve this problem in an industrialized future society, but the speed of climate change is likely to be substantially higher than the speed of development. Moreover, in the

natural world this solution is unavailable and both plants and animal life will suffer even more than humans. Just as an example from the Stern report (Table 6.1), peanut plants in India gave around 50 seeds per plant at temperatures up to 33.5°C during the flowering season. At 6°C higher values the yield was 10 times lower. It will be an interesting question whether genetic engineering will be able to increase the heat tolerance of plants. The general yields in agriculture are expected to significantly diminish in much of Africa even for modest temperature increases. Rising temperatures, unfortunately, do not pose a problem to the many tropical microbes or their vectors, as malaria, cholera and others will thrive in the warmer climate.

The glaciers in the Himalayas will increasingly melt. Initially this may lead to increased water availability, with however the risk of catastrophic flooding through sudden drainage of ice-dammed lakes. In a few decades the glaciers will be greatly diminished and the river flow during the dry season will be reduced in the great rivers of South-East Asia with damaging effects on agriculture in the valleys (e.g. the Ganges valley) which feed hundreds of millions of people. Similar problems arise in the Andean areas of Latin America, and the glaciers in central Africa will have gone even sooner. Rising sea levels will have catastrophic effects in the huge agricultural deltas in South-East Asia. Not only will some land become fully covered by the sea, but storm floods will reach much further inland and salinate the soils. Of course, these problems will not be restricted to countries in the tropics such as Bangladesh, but the Nile delta, which will soon host 100 million people, and many islands and low-lying countries such as Holland may gradually become uninhabitable. Again in the developed world engineering solutions may well allow a sea level rise of a meter to be accommodated, but even there an extra 5 meters, as occurred during the last interglacial, will stretch the possibilities.

Droughts have brought whole civilizations to ruin, even during the relatively stable climate of the Holocene. Examples include the Mayan empire, the Indus valley society and others. More recently, since the 1960s, the Sahel region has suffered from a catastrophic drought. Climate models show a major drying in several already dry subtropical areas which risk changing from dry to more nearly desertic. The Mediterranean region and the southwest USA/northern Mexico areas are examples, with potentially serious losses in agriculture.

In the north, the melting of the permafrost may pose the greatest risk. In much of Alaska, northern Canada and Siberia, buildings have been constructed on the solid permafrost, but with the rapid melting a complete reconstruction may be needed.

6.7 Appendix

6.7.1 The four main SRES scenarios

In these scenarios the IPCC has attempted to construct the future world population, CO_2 emissions, energy use, GDP, etc., in a coherent way on the basis

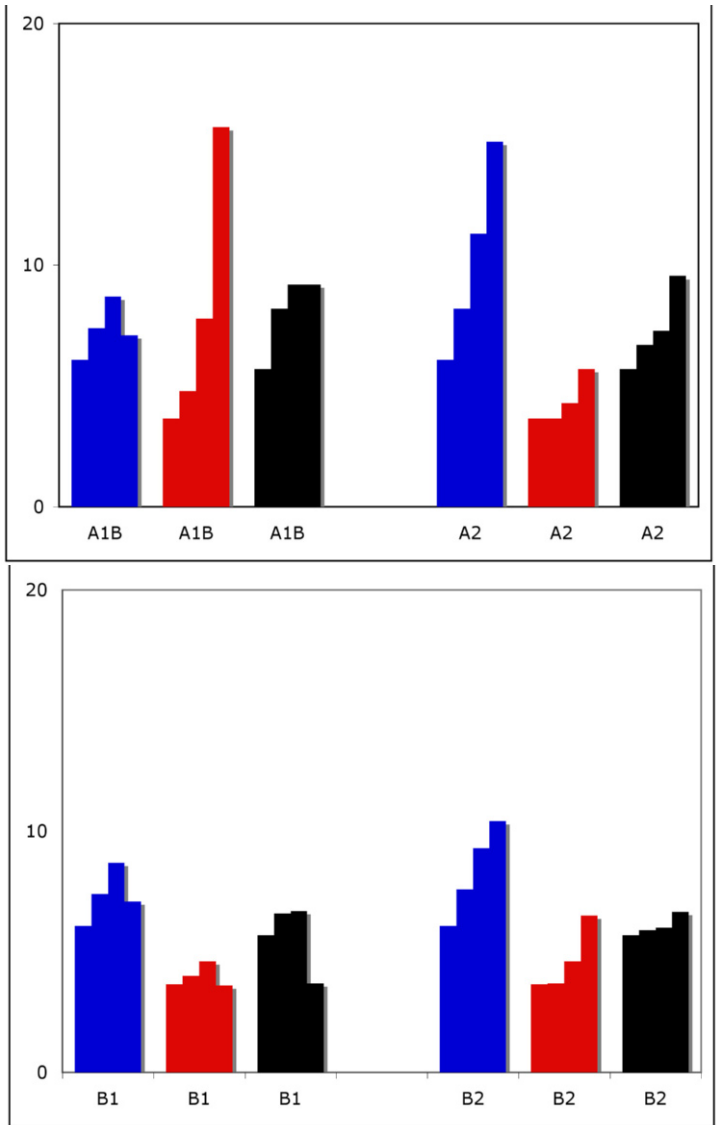


Figure 6.5 Population, annual energy use and CO_2 production for four IPCC scenarios.

of different assumptions about humanity's priorities (see earlier discussion in this chapter). Figure 6.5 shows for each of the scenarios A1B, A2, B1 and B2 in blue, the population in thousand millions, in red the per capita annual use of primary energy in units of 100 gigajoules, and in black the annual per capita CO_2 production in units of 0.2 ton of carbon. In each case the data are given from left to right for the years 2000, 2020, 2050 and 2100.

In the A2 scenario the per capita energy use is relatively low, but the low

development leads to a rapid population growth and a rather high production of CO_2 per unit of energy. As a result, the total CO_2 production is very large. Scenario B2 follows the UN population projections. Both B scenarios have relatively modest energy use, in contrast to A1B, which is a fast development scenario with much use of renewables.

6.8 Notes and references

[1] Ausubel, J.H., 1991, 'Does climate still matter?', *Nature* **350**, 649–652.
[2] IPCC: AR4-WGII, SPM, p. 17.
[3] IPCC: AR4-WGI, SPM.
[4] IPCC: *Third Assessment Report* (TAR), WGI.
[5] IPCC: *Special Report on Emission Scenarios* (SRES).
[6] Stated in caption to Figure 1 of IPCC–SRES, SPM-1, 2000.
[7] The values in Table 6.2 are based on the averaging of many climate models. In the SRES scenarios the calculated CO_2 concentrations in different climate models range over +28% to –11% of the average values. In the stabilization scenarios the calculated carbon emissions range from +16% to –26%. The ΔT values in the different models cover the range from +63% to –43%. The CO_2 concentrations are from the IPCC: TAR-WGI, as are the emissions, except for WRE 450 and 1,000 which are from IPCC: AR4-WGI. The first four ΔT's are also from this last report, the last three from the TAR. In some reports the emissions are given as tons of CO_2 with 1 ton of carbon corresponding to 12/44 tons of CO_2.
[8] Wigley, T.M.L et al., 1996. 'Economic and environmental choices in the stabilisation of atmospheric CO_2 concentrations', *Nature* **379**, 242-245.
[9] Sutton, R.T. et al., 2007, 'Land/sea warming ratio in response to climate change: IPCC AR4 model results and comparison with observations', *Geophysical Research Letters* **34**, L02701, 1–5.
[10] Winton, M., 2006, 'Amplified Arctic climate change: What does surface albedo feedback have to do with it?', *Geophysical Research Letters* **33**, L03701, 1–4.
[11] Precipitation projections from Giorgi, F. and Bi, X., 2005, 'Updated regional precipitations and temperature changes for the 21st century from ensembles of recent AOGCM simulations', *Geophysical Research Letters* **32**, L21715, 1–4. These are for the period 2070–2099 with respect to 1960–1979 under the A1B scenario. Temperature projections are from Giorgi, F., 2006, 'Climate change hot-spots', *Geophysical Research Letters* **33**, L08707, 1–4, for the period 2080–2099 with respect to 1960–1979 as an average for the scenarios A1B, A2 and B1. The factors relative to global warming should not be too sensitive to the particular scenario and all these values for temperature and precipitation should differ from those for 2100 with respect to 2000 under the A1B scenario by much less than their still large uncertainty.

[12] Seager, R. et al., 2007, 'Model projections of an imminent transition to a more arid climate in southwestern North America', *Science* **316**, 1181–1184.
[13] Liu, Z. et al., 2006, 'On the cause of abrupt vegetation collapse in North Africa during the Holocene: Climate variability vs. vegetation feedback', *Geophysical Research Letters* **33**, L22709, 1–6.
[14] Luterbacher, J. et al., 2004, 'European seasonal and annual temperature variability, trends, and extremes since 1500', *Science* **303**, 1499–1503.
[15] Lambeck, K.M. et al., 2004, 'Sea level in Roman time in the Central Mediterranean and implications for recent change', *Earth and Planetary Science Letters* **224**, 563–575.
[16] IPCC-AR4-WGI, SPM, p. 7.
[17] Alley, R.B. et al., 2005, 'Ice-sheet and sea level changes', *Science* **310**, 456–460.
[18] Vaughan, D.G. and Arthern, R., 2007, 'Why is it hard to predict the future of ice sheets?', *Science* **315**, 1503–1504.
[19] Thomas, R. et al., 2006, 'Progressive increase in ice loss from Greenland', *Geophysical Research Letters* **33**, L10503.
[20] Stearns, L.A. and Hamilton, G.S., 2007, 'Rapid volume loss from two East Greenland outlet glaciers quantified using repeat stereo satellite imagery', *Geophysical Research Letters* **34**, L05503.
[21] Ekström, G. et al., 2006, 'Seasonality and increasing frequency of Greenland glacial earthquakes', *Science* **311**, 1756–1758.
[22] Velicogna, I. and Wahr, J., 2006, 'Acceleration of Greenland ice mass loss in spring 2004', *Nature* **443**, 329–331.
[23] Horwath, M. and Dietrich, R., 2006, 'Errors of regional mass variations inferred from GRACE monthly solutions', *Geophysical Research Letters* **33**, L07502.
[24] Overpeck, J.T. et al., 2006, 'Paleoclimatic evidence for future ice-sheet instability and rapid sea-level rise', *Science* **311**, 1747–1750.
[25] Otto-Bliesner, B.L. et al., 2006, 'Simulating Arctic climate warmth and icefield retreat in the last interglaciation', *Science* **311**, 1751–1753.
[26] Comiso, J.C., 2006, *Geophysical Research Letters* **33**, L18504, 1–5.
[27] Gregory, J.M. et al., 2004, 'Threatened loss of the Greenland ice-sheet', *Nature* **428**, 616.
[28] Oppenheimer, M., 1998, 'Global warming and the stability of the West Antarctic Ice Sheet', *Nature* **392**, 325–332, contains a useful map identifying Antarctic features.
[29] Rott, H. et al., 1996, 'Rapid collapse of Northern Larsen ice shelf, Antarctica', *Science* **271**, 788–792.
[30] Scherer, R.P. et al., 1998, 'Pleistocene collapse of the West Antarctic Ice Sheet', *Science* **281**, 82–84.
[31] Domack, E. et al., 2005, 'Stability of the Larsen B ice shelf on the Antarctic Peninsula during the Holocene epoch', *Nature* **436**, 681–685.
[32] Crucifix, M. et al., 2006, 'The climate response to the astronomical forcing', *Space Science Review* **125**, 213–226.

[33] Pagani, M. et al., 2005, 'Marked decline in atmospheric carbon dioxide concentrations during the Paleocene', *Science* **309**, 600–602.
[34] Zachos, J.C. et al., 2005, 'Rapid acidification of the Ocean during the Paleocene-Eocene thermal maximum', *Science* **308**, 1611–1615.
[35] Rahmstorf, S.A. et al., 2007, 'Recent climate observations compared to projections', *Science* **316**, 709.

7 The Future of Survivability: Energy and Inorganic Resources

It would seem to be a fact that the remotest parts of the world are the richest in minerals and produce the finest specimens of both animal and vegetable life.

Herodotus

7.1 Energy for 100,000 years

An ample supply of energy is an essential requirement for the continuation of our civilization. We need energy to heat or cool our houses, to move cars, trains and aircraft, to power our machines and computers and to run our chemical industry. Current energy production comes mainly from oil, natural gas and coal, which are derived from biomass accumulated over many millions of years, and thus represent past solar energy buried underground. The solar energy ultimately comes from the nuclear fusion reactions that take place in the hot interior of the Sun and convert hydrogen into helium. Not surprisingly, attempts are being made to extract energy from the same reactions on Earth, but it has not yet been possible to confine the hot gas in the small volume of a reactor. However, nuclear reactors based on radioactive uranium have been successful and contribute modestly to present-day energy supplies. Again, ultimately the energy comes from a celestial source, the violent supernova events (exploding stars) during which the uranium has been synthesized, which was later incorporated into our Solar System.

The Sun warms the oceans, evaporating some of the water. The resulting clouds drift inland where their rain may fall on high ground. From the resulting downward flowing rivers and streams, hydroelectric power may be extracted, which is an important energy source in several countries. The Sun heats the Earth very nonuniformly: the equator receives the majority of the Sun's heat, and the polar regions receive very little. This creates winds and for many centuries wind mills have been built to tap some of this wind energy. The solar energy also makes plants and trees grow, and burning the resulting biomass has not only been an important source of energy in the past, but still is for many people in the less-developed countries. Currently attempts are being made to convert biomass into biofuels. The efficiency with which plants use solar energy is low, usually well below 1%. Much more energy may be obtained directly from

the Sun by the absorption of its radiation on dark surfaces for heating or on solar cells for electricity production, but to date the direct collection of solar energy has made a negligible contribution to our power supplies, partly because of technological problems, and partly because of insufficient motivation as long as oil, gas and coal are not too expensive.

Minor sources of energy include geothermal energy where the internal heat of the interior of the Earth is tapped. The energy of the tides and of waves in the oceans may also make a contribution. The tides result primarily from the gravitational attraction exerted by the Moon, and their energy is dissipated by friction in the oceans. Finally there is the energy associated with the osmotic pressure that results when fresh water meets saline water, or when warmer and colder water are brought together in the oceans. The main problem with all of these geo-energies is that they are very diffuse in most places on Earth.

There has been much discussion in recent years about the dematerialization of the economy and also of its de-carbonization [1]. What this means is that the amount of steel or the amount of energy per unit of GDP (Gross Domestic Product in units of dollars, euros or other) is decreasing. While that may be very satisfactory, it is not very illuminating in terms of resource utilization since, in many cases, the absolute amount of resources used does not diminish. In fact, almost everywhere the amount of energy used per capita is still increasing, and so does the consumption of many material resources. GDP helps little towards heating our houses in winter. Instead this takes a certain amount of energy, and while it may be very gratifying that it costs a smaller fraction of our income, this does not change the problems associated with the insufficiency of oil or gas or those relating to the CO_2 emissions. Of course, it is different if we take such measures as switching to renewables or utilizing energy in more efficient ways. It has been frequently pointed out that current automobiles are very inefficient, with tank to wheel efficiencies of 16% or less [2]. So 84% of the energy in the gasoline in the tank is wasted, and as only 16% is really used for moving the car against air and road resistance, a switch to partly electrical hybrid cars could probably double the energy efficiency. Economies are also possible in industrial processes, in the heating of buildings and in many other areas.

Many projections have been made of energy use in the coming century, but such projections depend on uncertain models of economic growth, technological developments and resource availability. In some respects it is easier to foresee the energy supply needed to maintain the 100,000-year society. There may be much doubt about the future availability of oil, but we may be sure that rather early in the 100,000 years we will have reached the end of accessible hydrocarbons or perhaps more likely that we would not dare to use them because of their impact on the climate. We may also be uncertain about the speed of development in different parts of the world. But if the premise of Chapter 1 is correct – that a certain level of equality is required for the long-term survival of our civilization – then on the 100,000-year timescale the result will not depend very much on whether the less-developed countries arrive there in 50 years or in two centuries. So we begin by considering the long-term situation.

7.1.1 Energy requirements for the 100,000-year world

Estimates of the future energy requirements of the world are notoriously difficult to make and frequently have been erroneous even on a timescale of a few decades. As an example, 17 projections made during 1969–1973 for the (primary) energy supply needed in the year 2000 varied between 500 and 1,400 EJ [3]. The actual value turned out to be slightly less than 400 EJ! In Chapter 6 we have discussed the scenarios created for the IPCC which, for the year 2100, project energy needs in the range 500 to more than 2,000 EJ. Such projections are based on different assumed rates of population growth, GDP and energy intensity (energy/GDP). Somewhat further into the future the uncertainties can only increase. (The units of energy are given in Box 7.1.)

Before proceeding further we should briefly note the difficulties that arise when both heat energy and electrical energy are to be added in a global balance.

Box 7.1 Units

Electrical and mechanical energy are expressed in joules. If such energy is converted into heat energy, 1 joule corresponds to 0.24 calories, with 1 calory the energy needed to heat 1 gram of water by 1°C. An important fact is to be noted: electrical and mechanical energy are entirely convertible into heat energy, but the inverse is not the case. In an electrical generator typically only a third or half of the heat is converted into electrical energy, and the remainder leaves the generator as 'waste heat' into the air or in cooling water. So electrical energy represents a higher quality energy than heat energy. Electrical energy flow is measured in watts: 1 watt = 1 joule per second. Frequently larger units are needed which are given as powers of 1,000 as follows: kilo, mega, giga, tera, peta, exa for 10^3, 10^6, 10^9, 10^{12}, 10^{15}, 10^{18}. Only the first letter is used, e.g. 1 MW = 1 million watts, etc. Energy flows are measured in watts, energies by multiplication with the duration of the energy flow. The most often used unit is the kilowatt-hour (equal to 3,600,000 joules), a unit particularly suited for household purposes. For discussing global energy problems the exa joule (EJ) is more appropriate (equal to 278 billion kWh or 278 TWh) or alternatively the terawatt year (TWyr) equal to 31.6 EJ. The EJ is convenient when a mix of energies is to be discussed, the TWh or TWyr when an all electrical world is considered.

The quantity of oil is usually expressed in barrels with 1 barrel = 159 liters, or in tons (1 ton = 7.35 barrels), with 1 Gt of oil having an energy content of 42 EJ. Natural gas is measured in m^3 at atmospheric pressure with 1,000 Gm^3 containing an energy of 38 EJ. The energy content of coal is slightly more variable with 1 Gt containing around 20–25 EJ. Current world supply of primary energy corresponds to some 500 EJ, increasing by 2.1% per year; this figure includes an uncertain 36 EJ of traditional biomass energy in the developing countries.

Much of present-day electricity is produced in power plants in which the steam or hot gas that results from the burning of fossil fuels or from nuclear processes is used to drive a turbine. After the gas comes out of the turbine it is cooled, so that the pressure on the intake is higher than that at the exit; the pressure difference then drives the turbine which may be attached to a dynamo to generate electricity. From fundamental thermodynamics it follows that there is a maximum conversion efficiency that is more favorable the higher the temperature of the gas before the turbine and the cooler after it. In practice, the efficiency has generally been no more than one-third, although in more recent power plants values around one half have been obtained. In official statistics, more or less by definition, the efficiency of nuclear power plants sometimes has been set at 1/3. In a hydroelectric plant the mechanical energy is converted into electrical energy, a process that in theory can take place at 100% efficiency. This leads to the paradoxical result that, usually, hydroelectricity counts for less in the statistics than nuclear energy. In fact, in the present-day world hydro and nuclear electricity are about equal in terms of numbers of kWh produced per year but, in primary energy, nuclear is counted as three times greater. In the language of the International Energy Agency (IEA) 'total primary energy supply' includes all energies that enter the world's energy system, while 'total final consumption' includes the electricity consumed, but not the energy lost as waste energy in the production of that electricity. Much of the energy in the 100,000-year world is likely to be in the form of electricity: hydro, wind, solar and the output of nuclear or fusion reactors.

In the following we shall express all energies in electrical units, although this does not solve all our problems. For example, solar electricity involves various efficiency factors: duration of sunshine, efficiency of solar cells, losses in electrical cables when produced far from the user, etc. These have then to be taken explicitly into account.

To assess the energy needs of the 100,000-year world we make two assumptions. For the population we take 11 billion people, which is the medium stabilization level from the UN projections. For the energy consumption we assume that the per capita value will correspond to the average in 2002 of that in the USA and in the more advanced European countries. The year 2002 has been chosen because thereafter the energy markets have been shaken by general turbulence, the origin of which may in part be related to speculation and in part to political factors. There is some evidence that, in fact, a plateau has been reached in the energy consumption of the economically more advanced countries. As an example, in the USA and Canada the primary energy supply per capita from 1971 onwards increased by 0.1% per year, while the more relevant 'total final consumption' (based on data of the International Energy Agency) declined by 0.3% per capita per year over the same period. We then find the 'final energy consumption', the energy actually consumed by the end user per year in electricity, hydrocarbons, heat and renewables, to be 0.23 EJ per million people in the USA and 0.13 EJ in France, Germany, the UK, Belgium and Holland, for an average of 0.18 EJ. With our assumption of 11 billion people at

more or less the same level of well being, this then corresponds to about 2,000 EJ per year. This is about a factor of 7 above the estimated world energy consumption level for 2002.

Why is there such a large difference between the USA and Europe? A particular factor is transportation. As an example, the French in 2003 used 61% less energy for this (per capita) than the Americans. One only has to look at the big cars on a US highway to see the reason. Of course it is also true that in a less densely settled country transportation is likely to be somewhat more expensive. Furthermore, the USA has been accustomed to cheap energy and, consequently, has been much more wasteful than the EU where additionally high taxes on oil have had a beneficial effect on keeping consumption at a lower level. Of course the figure of 2,000 EJ of electrical power annually is very uncertain. On the one hand, present-day energy use remains rather wasteful and so economies are certainly possible. On the other, it is rather clear that the exploitation of very much poorer ores to extract needed minerals, and the need for desalinated water, will increase energy consumption. The main importance is that it gives us a yardstick by which to measure the options for different kinds of energy supplies. The figure of 2,000 EJ corresponds to an energy flow of 63 TW or 63 TWyr annually.

We shall now inspect the different contributions that may be obtained from various sources, as summarized in Table 7.1. In Section 7.1.2 we shall discuss the three minor sources and in Sections 7.1.3–7.1.8 the six major ones. For each of the six we shall evaluate the consequences from the assumption that they contribute equally to the total energy mix, i.e. 10 TW each. Of course it is very well possible that in the long run some will be favored over others.

Table 7.1 Possible sources of power for the 100,000-year world

Source	Potential	Problems
Geothermal	Probably minor	Diffuse
Ocean tides and waves	Relatively minor	Diffuse
Hydroelectricity	Minor	Water, environment
Wind	Important	(Environment)
Solar photovoltaic	Large	Necessary materials
Solar thermal	Large	
Biomass	Important	Competition for land
Nuclear	Large	Thorium, avoid plutonium
Fusion	Large	To be demonstrated

7.1.2 Minor energy sources for the long-term future

Geothermal energy

The heat in the interior of the Earth derives from the gravitational energy liberated during its formation and from the radioactive decay of uranium, thorium and an isotope of potassium (^{40}K) in the Earth's crust. The total heat

energy of the hot rocks below the Earth's surface is very large and so it is not surprising that over the last century the technology has been developed to extract hot water and electricity. At a depth of 5 km the temperature reaches, on average, some 150°C. Rain water may make its way to such depths and generate steam. By drilling into steam reservoirs we may drive a turbine using the high-pressure steam to generate electricity. Alternatively, in dryer areas we may inject cold water and recuperate hot water or steam that can be used for heating homes and greenhouses. The energy flow of the heat towards the surface has been measured in thousands of places. Over the whole surface of the Earth it has been found to amount to 44 TW of heat energy [4]. Over the land it amounts to some 10 TW. It is very diffuse and mainly useful in volcanic areas where it is more concentrated or in areas of active tectonics. At present, according to IEA figures, geothermal electricity production world wide is no more than 0.008 TW – small even by the standards of renewables. About twice as much heat energy is obtained as hot water. According to a report by the International Geothermal Association, the global electricity production could ultimately reach some 2.5 TW [5], corresponding to some 4% of the need in the future long-term society. So geothermal energy may make a useful contribution, but it is unlikely to become a major global source. Much of the oceanic heat flow comes from the ridges where new continental crust forms (Section 2.4). For the moment drilling into these mid-ocean ridges would seem to be a horrendous undertaking.

Ocean tides and waves

The tides are caused by the difference in the gravitational forces due to the Moon – and to a lesser degree the Sun – on the oceans and on the Earth as a whole. This causes the ocean surface to rise and fall by a few decimeters. The resulting motions are dissipated by turbulence in the deep ocean and by friction on the ocean bottom. When the tidal bulge reaches coasts or bays the water is pushed up and reaches greater heights – in the case of the Bay of Fundy in Nova Scotia by up to 15 meters. In such places the motion of the water can be used to drive a turbine and generate electrical power. The global energy flux through the tides is some 3.6 TW [6], but in most places it is too diffuse for practical power generation. Also wind-driven waves may be used. Currently, tidal and wave electrical power generation is even 20 times less than that produced geothermally.

In the 1960s there were many ideas to utilize the difference in temperature between the surface of the ocean and that deeper down, or differences in salinity, in estuaries to produce electrical power [7]. Owing to the diffuse nature of all of these, the results have been negligible. Evidently the use of oceanic energies requires robust equipment in order to withstand the storms or cyclones that may occur. Even though it is difficult to specify hard upper limits on the energy to be gained from the geological processes, other power sources seem a lot easier to realize.

Hydroelectrical power

This is indirectly based on the solar energy that evaporates water from the oceans. The water vapor is transported to altitudes of hundreds or thousands of

meters in the atmosphere, drifts inland and upon condensing produces rain. The resulting rivers flow downhill and some of their energy may be converted into electrical energy. A crude estimate of the maximum hydroelectric potential is easily made. The world's rivers annually transport 40,000 km^3 of water to the oceans. The mean altitude of the Earth's land is 860 meters. Some of the rain will fall above this altitude and some below. If, for simplicity, but too favorably, we assume that all rain falls at the mean altitude, the total energy flux of the water on the way down would correspond to 12 TW [8]. In natural circumstances much of the energy is dissipated by friction in the riverbeds.

However, various circumstances limit the hydroelectric potential. While hydroelectric plants are intrinsically clean without producing CO_2 and other pollutants, except during the fabrication of the steel for turbines and tubes, they have serious environmental impacts. In higher mountain areas these are limited to the loss of mountain streams. When much of the rainfall is conducted through the necessary pipes, an arid landscape remains. In regions with less steep slopes, large reservoirs behind high dams are needed which implies the flooding of large areas of land. This may not be too disastrous in desert areas, such as around the Assouan dam in the Nile, but in more densely populated areas the problem would be serious. A recent example is the Three Gorges Dam in the Yangtse river [9]. Here some 18 GW (0.018 TW) of electrical power was to be obtained, but nearly two million people had to be displaced because of flooding behind the dam. This project had other hydrological reasons, but it illustrates the problems that occur. It has been found that the maximum technically feasible global hydropower production would be about 15,000 TWh per year (1.7 TW), less than 3% of the total future energy requirement [10]. Even that amount is likely to do a great deal of ecological damage. Most of the growth of hydropower during the first three decades of this century is expected to occur in the less-developed countries.

Even though they could make valuable contributions, the total of the geological sources discussed so far is not likely to exceed 5% of the total requirement for the 100,000-year society. So we next turn to the six more promising energy sources: wind, solar photovoltaic, solar thermal, biomass, nuclear and fusion. As we stated above, we shall for the moment adopt a model in which each of the six contributes 10 TW to the energy requirement.

7.1.3 Wind energy

Windmills in previous centuries had low efficiencies, but an increased understanding of the air flows around the blades has led to improved designs for high-performance wind turbines. These have typically rotating blades some 40 meters long on an axis mounted 80 meters above ground where the wind is stronger than at the surface. The maximum power rating is typically 1,500 kW. Of course the wind is not blowing continuously and in practice the mean output is no more than some 30% of the maximum. Under favorable conditions a turbine rated at 1.5 MW could then generate some 5 million kWh of electricity per year. More recently an even larger installation has been constructed, rated at 4,500 kW.

Early wind turbines were expensive and suffered from frequent breakdowns, but further industrial development has changed this. The cost of wind energy has come down from more than 50 cents per kWh in 1980 to 4–7 cents per kWh today [11]. (Here and elsewhere, we adopt US cents, unless another unit is specified.) With much disagreement about whether this is more or less expensive than the energy from coal when all costs are included, it is clear that wind energy is now based on proven technology, and it is affordable. There are, however, some negative aspects. Some people object to the wind turbines because they find them unesthetic. Have they ever looked at coal-fired generators and at the pollution they cause? In addition, wind turbines tended to be noisy, which certainly has to be taken into account in their siting, though recent advances have greatly reduced the noise.

It has been estimated that the total wind power that theoretically could be tapped would be some 70 TW, more than the future total energy requirement [12], although practical considerations may very much reduce this. Some of the best sites are in such places as the rim of Antarctica, islands in the southern ocean, the Aleutian Islands, etc., but are not connected to an electrical net. Of course, even if it were not feasible to transport the electricity, one could produce hydrogen *in situ* by electrolysis of water and move this by ship. Other particularly favorable areas include the coasts of northwestern Europe and northern North America, the US mid-west and the great lakes region, Patagonia and some coastal areas of Australia (Figure 7.1). Unfortunately, the tropical countries and China have relatively few good sites. However, even in regions of low mean windspeed, more localized suitable areas may often be found. Suppose we would wish to obtain 10 TW from wind that is about one-sixth of the long-term electricity requirement. Taking into account that the power production is less than the power rating, we would need over six million wind turbines with a 5-MW power rating. To avoid interference between them there should not be more than six of these per 100 ha of land, corresponding to a total land requirement of around 100 million hectares, or less than 1% of the Earth's land area. Of course, that land could also be used for other purposes like agriculture or solar energy installations, since the physical space required for the wind mills is small. However, it is more likely that many wind farms would be placed in shallow seas, where the wind is stronger than on land and where esthetics are less of a problem. However, some measures might then be necessary to protect sea birds.

A detailed study of offshore wind turbines along the US east coast from 34 to 43°N, which takes into account that some areas are excluded for natural and navigational reasons, shows that by placing these out to a depth of 100 meters there is a wind potential of 0.33 TW [13]. Hence the total North American coastal potential would probably be of the order of 1 TW. It seems, therefore, not at all outlandish to believe that a total worldwide potential of 10 TW could be realized.

The main problem with wind energy is its intermittency: when the wind stops, the power stops. However, if a number of wind farms separated by

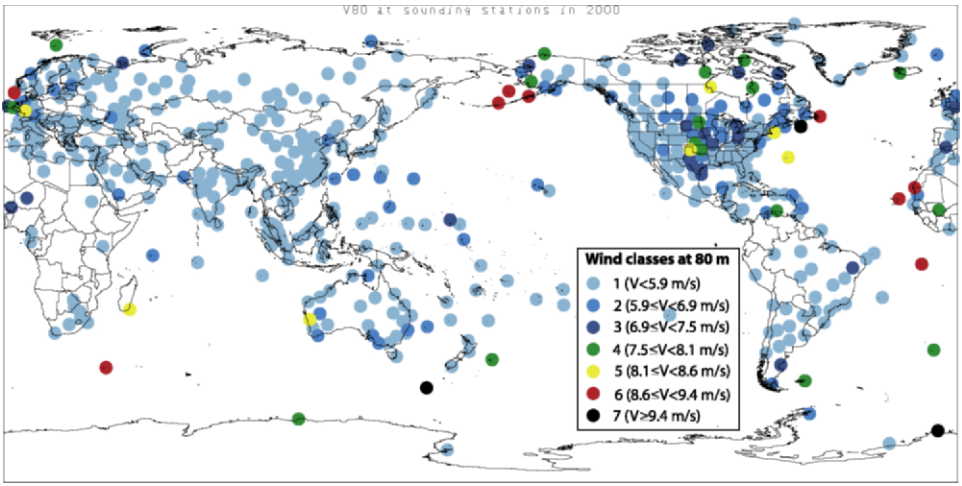


Figure 7.1 The global distribution of windspeeds at 80 meters height. Sites with wind classes 3 and higher are suitable for wind energy generation. (Courtesy *Journal of Geophysical Research* 110, D 121110, pp. 1–20, 2005, 'Evaluation of global wind power' by C.L. Archer and M.Z. Jacobson, Figure 2.)

hundreds of kilometers are feeding into a common electricity grid, it could be expected that the fluctuations would be much reduced [13]. Also, if the electricity is used to produce hydrogen, an occasional interruption would not be too serious. Nevertheless, it may well be that a 10–20% contribution from wind energy is as much as realistically can be considered.

7.1.4 Solar energy

Taking into account that some 30% of solar radiation is reflected back into space and considering only the part that falls on land, we obtain an energy flow of 35,000 TW, some 500 times larger than the required electrical energy for our long-term society. There are two ways in which solar energy may be transformed into electricity: by photovoltaic cells or by turbines driven by solar heat.

Solar photovoltaic cells

These are devices in which sunlight ejects electrons from sensitive materials and ultimately leads to the generation of electricity. Such 'photovoltaic cells' have achieved efficiencies of up to 41% in the laboratory and more typically of 10% with industrially produced cells [14]. At 15% efficiency in the subtropical deserts, some 30 million hectares of cells would be needed (less than 30 m^2 per person in the world) to generate 10 TW. Since additional space is needed for various purposes (access, transformers, etc.), a total of some 50 million hectares would be required, equal to the area of France. In the hot deserts of the world more than a billion hectares of land would be available. However, at present the cost is high, around 5–10 times that of wind energy, but further development and industrial

mass production should bring the cost down. One only has to remember the early CCD chips for imaging which were priced well above €1,000, while now every €100 camera has a CCD chip with far superior performance.

High-efficiency solar cells frequently use rare metals with special properties. As an example, a cell composed of layers of a gallium–indium–phosphor compound with a total thickness of nearly 0.001 mm [14] has been described with 30% efficiency, under concentrated light. To cover 30 million hectares of photocells, more than 1 million tons of the rather rare element indium would be needed, although this might be reduced if the light could be concentrated. However, current world indium reserves have been evaluated at only 6,000 tons. While ultimately perhaps more indium will be found, the difference between these figures is very large indeed. So apart from efficiency, the use of more common materials has a fundamental importance.

As also with wind energy, there are serious problems with intermittency and therefore efficient storage of the energy is needed. In suitable regions, this could take the form of water reservoirs in the mountains: when the Sun shines some of the energy could be used to pump water up, and at other times hydroelectric energy could be generated. Alternatively, the electricity could be used to dissociate water into hydrogen and oxygen. The hydrogen could be stored or transported and could later be used to generate electricity. Such processes, however, entail losses, and geographical distribution also causes problems. Dry deserts tend to be unpopulated and directly transporting electricity through high-tension lines also causes losses. But perhaps a future society would have the good sense to place the energy-intensive industries at the sites of the energy – as is already done today with part of the energy-intensive aluminum production that is now being located in areas of high hydroelectric potential, like Iceland. Sun and wind have a certain complementarity: maximum Sun is in the subtropics and maximum steady wind is at higher latitudes.

Solar thermal energy

This could also be used directly to heat water. On a small scale this may be done by having water flow below a dark surface exposed to the Sun. In fact, this has been successful in generating hot water for home heating, but to produce electricity efficiently, high temperatures are required for which a concentration of the solar energy is needed, which could be achieved by a system of mirrors that focus the solar light to a much smaller area. Other systems are based on creating a sort of greenhouse with hot air and extracting energy by letting the air move out through a turbine [15]. It remains to be seen what efficiencies can be reached; they should be no less than the 15% we conservatively assumed for the solar cells and the areas needed to collect the solar energy should therefore not be too different. Which of the technologies is preferable can only be decided on the basis of experience. What will be the real efficiency of the two approaches and what will be their cost? Of course, the problem of intermittency remains: what to do during cloudy days? Hence, a storage facility is required. Similar to wind energy, this may limit solar energy to 10–20% of the total energy supply.

7.1.5 Biofuels

During most of history humanity warmed itself by burning wood and clothed itself with materials of contemporary biological origin. Even today traditional biomass (mainly firewood and other organic matter) still accounts for 7% of the world's primary energy. Unfortunately it is used inefficiently. By the middle of the last century fossil fuels had largely taken over not only for providing heat and locomotion, but also as a source of organic materials like plastics and other synthetics. The consequence has been an excessive production of CO_2 and the resulting global warming.

Plants synthesize organic materials (cellulose, sugars, starch, etc.) from atmospheric CO_2 and minerals in the soil by photosynthesis – the process of using the energy of the Sun's light. When the plants die and rot away they return the CO_2 to the atmosphere. If we use plants to make biofuels, we consume in principle as much CO_2 as is produced when we burn that fuel. While this may seem an ideal solution to the energy/climate problem, there are, of course, problems relating to agricultural land, water and fertilizer in addition to the technological difficulties in efficiently converting plant material into fuel. That conversion may require energy from fossil fuels and so a careful analysis is needed to determine the CO_2 balance. Since technological developments are taking place rapidly, it is difficult at present to evaluate the ultimate possibilities.

One of the main problems is that photosynthesis is not a very efficient process: only 0.1 to 1% of the solar energy is converted into usable plant energy. In this respect solar cells with efficiencies of 10% and more are superior, with the important corollary that much less land is needed to generate a given amount of energy. Moreover, solar cells do not require agricultural land or water, and so a desert is an acceptable location that has also the advantage of, usually, maximizing the annual amount of sunlight. However, the positive aspect of biofuels is that they represent a minor modification to the present economy without all the complexities of the hydrogen economy or of electricity for road transportation. The intermittency problems due to cloudy days do not occur, since plants integrate the solar energy over the growing season.

In order to see what would be required to obtain the equivalent of 10 TW of energy (316 EJ per year) in the form of biofuel, we note that reported ethanol or other biofuel annual productivity amounts to typical values in the range of 3,000–7,000 liters per hectare [16, 17]. So, as an average, we shall adopt an annual productivity of 5,000 liters per hectare of ethanol, a little more than we would obtain if we process corn to make the ethanol and a little less than we could obtain from sugar cane. This corresponds to about 100 GJ per hectare of energy, from which it follows that some 3,000 million hectares of land are required for one harvest per year. This may be compared with the 680 million hectares currently devoted to the world's cereal production (Section 8.2). It is equal to 20% of the Earth's land area!

This estimate is still incomplete because it neglects the energy that is needed (a) to clear the land where the plant material is to be grown, (b) to cultivate the plants, including fertilizers and insecticides, (c) to harvest the plant material and

(d) to process it into ethanol in a biorefinery. Detailed calculations have been made to estimate all these energy outputs, and the results are disturbing. In the case of ethanol from corn it was found in different studies that the energy inputs totaled 29% more [18] or 5–26% less [19] than the energy output in the form of ethanol. Since fossil fuel is used not only in the refinery, but in the production of fertilizer, in transport and in other steps, the conclusion was that the greenhouse gases produced amounted to only 13% less than if one had used gasoline to obtain the energy. The rational conclusion of all of this is that the gains in corn biofuel are negligible. If one also takes into account the soil erosion and the excess fertilizer that arrives in the environment, the current frenzy in the USA towards corn ethanol is hard to understand as anything other than an agricultural subsidy program.

Some confusion could be caused by the article presented in reference [19], in which it is shown that the input of energy in the form of petroleum is in the range of only 5–20% of the energy produced as ethanol from corn. However, a much larger energy input comes from coal and natural gas, and this is the reason that the reduction of greenhouse gases is so insignificant. Therefore, this is not net energy from biomass, but rather the utilization of it to transform coal and gas into fuel for cars. It could be argued that the chemical industry might achieve the same result in various coal liquefaction processes without the ecological problems of the biofuels. The situation could be different in the future if efficient procedures would become available to convert cellulosic material into ethanol.

The situation is more favorable if sugar cane is used, as has been pioneered in Brazil. Since sugar is much easier to process in the biorefinery than corn starch, it is estimated that the resulting ethanol contains 10 times more energy than is put in as fossil fuel [20]. As a result, the Brazilian ethanol industry is operating without subsidy, while the USA has placed import duties on the Brazilian product to protect its subsidized corn product! In Europe sugar beets are beginning to be used, but the resulting ethanol contains only twice as much energy as the fossil fuel input and it is more than three times as expensive as the Brazilian product. The distortions of the markets by subsidies for agricultural products are also fully visible in the biofuels.

Recently, palm oil has gained in importance. It is a clear example of the dangers of the biofuels: some use it as cooking oil, others to make ethanol (or cosmetics!), and so there is direct competition between food and fuel. Unfortunately the same climatological circumstances that favor the growth of the rainforests are also optimal for the oil palms and, consequently, further contribute to tropical deforestation.

Much of the world's plant material is composed of cellulose, which is more difficult to process than sugar, and research is being done to see how the 'biomass recalcitrance' can be overcome [21]. If these efforts are successful the picture will change a great deal, with grasses becoming feedstock for biofuels – North American switchgrass and tropical African elephant grass are frequently mentioned. Also trees could be of interest if the lignins could be broken down

efficiently. Native grass land perennials could be grown on land unsuitable for agriculture and appear to be particularly effective when species diversity is maintained [22].

Finally there are the proposals to grow algae in water tanks, but such proposals are still at a very early stage of development [23]. However, all proposals for utilizing biomass suffer from the low efficiency of photosynthesis and the resulting land area requirements. Perhaps genetically engineered plants with superior performance will improve the situation, but it is essential to evaluate very carefully the ecological consequences. Above all, food production for 11 billion people should be the prime purpose of the agricultural world and should not be allowed to come in direct competition with biofuels. In this respect, recent price increases in agricultural commodities, which in part are due to conversion of agricultural land to corn for biofuel, are a worrisome presage for the future. A small area of photovoltaic cells placed in the desert is likely to have less of an ecological footprint than a farm for biofuels.

7.1.6 Nuclear energy

In an atomic nucleus the protons and neutrons (see Box 7.2) are much more tightly bound than the atoms in a molecule. As a consequence, in nuclear reactions typically a million times more energy is involved than in combustion or other chemical processes. The exploitation of nuclear energy began soon after the required nuclear physics had been understood. By now, in common usage, 'nuclear energy' refers to energy obtained from the fission of heavy elements and 'fusion energy' to that from the fusion of light elements. In practice then, nuclear energy involves uranium, thorium or their reaction products like plutonium.

After the construction of the first nuclear reactors in the early 1950s great optimism prevailed, symbolized by the phrase that 'nuclear energy would be too cheap to be metered'. By now some 450 reactors are in service which generate 2,800 TWh per year of electricity, some 15% of global electricity production. Three-quarters of these reactors were built more than two decades ago, and in several countries, including the USA and Germany, construction of new reactors has reached a standstill. The trauma of the Chernobyl accident in 1986 in which much radioactive material was released over parts of Europe, and the secrecy with which minor mishaps in reactors have been treated, have led to a loss of confidence in anything nuclear among large segments of the population. However, the realization that nuclear energy could contribute to reductions in CO_2 production from electricity generation has perhaps begun to create a more positive view. Nevertheless, any future important nuclear accident in the world could reverse this.

Natural uranium consists of two isotopes ^{238}U (99.3%) with a half-life ($t_{1/2}$) of 4,500 million years and ^{235}U (0.7%) with a $t_{1/2}$ of 700 million years. Both were produced more or less equally in supernova events before the Earth was formed, but by now much of the ^{235}U has decayed. ^{235}U has unique characteristics that make it a suitable fuel for a nuclear reactor. A low-energy (thermal) neutron

Box 7.2 Elements and isotopes

All natural matter on Earth is made of the atoms of 81 stable elements. In addition, there are two radioactive elements (uranium and thorium) with such long lifetimes that much of what existed when the Earth formed is still there. The atoms consist of a very compact nucleus surrounded by a cloud of electrons. The nucleus is composed of positively charged protons and uncharged neutrons of about equal mass. The number of electrons in an atom is equal to the number of protons in the nucleus. The number of neutrons is usually not very different from that of the protons in the lighter nuclei, but typically exceeds that by some 50% in the heavy nuclei.

The chemical characteristics of an element, the molecules they can form, are determined by the number and distribution of the electrons. Many elements have nuclei with different numbers of neutrons – called isotopes. Thus, hydrogen has three isotopes: ^{1}H with a one proton nucleus, heavy hydrogen, deuterium D or ^{2}H with one proton and one neutron, and tritium, T or ^{3}H with an extra neutron, but which is radioactive and decays with a half life of 12.5 years. All three can make the same molecules, like water H_2O by combination with an oxygen atom, though there are subtle differences in the speed with which they react.

The nuclei are much more tightly bound than the molecules, with the energies involved in nuclear reactions being typically a million times larger than those in chemical reactions. Reactions in which light nuclei fuse generally liberate energy, while the very heavy nuclei liberate energy by fission into less heavy ones. In nuclear processes a neutron may transform into a proton with the nucleus emitting an electron or the inverse with a positively charged electron – a positron – appearing. Also a high-energy photon (a quantum of light), a so-called gamma ray, may be emitted and in the case of radioactive decay also an alpha particle, i.e. a helium nucleus with two protons and two neutrons. Energetic electrons, gamma rays, neutrons and alphas constitute the much feared radiation of radioactivity. The alpha particles can be stopped by a sheet of paper but are dangerous upon ingestion, the electrons and gammas are more penetrating and protection from the neutrons requires a thick layer of lead.

causes it to fission and in the process additional, but more energetic, neutrons are produced. When these are slowed down they may cause additional ^{235}U nuclei to fission and so a chain reaction occurs. The slowing down of the neutrons may be achieved in suitable materials, such as water or graphite, that scatter the particles but do not absorb them. For example, in a reactor moderated by graphite rods all one has to do is to pull out the rods if the chain reaction becomes too strong or immerse them more deeply into the nuclear fuel if the reaction is too weak. So a condition of criticality may be maintained in which the

reaction proceeds at just a constant rate. Since the chain reaction cannot function when the concentration of ^{235}U is very low, the uranium has to be enriched to about 3–4%. This is now frequently done by chemically transforming uranium into a gas – uranium hexafluoride (UF_6) – and then placing this in very rapidly spinning centrifuges. The heavier $^{238}UF_6$ then experiences a slightly stronger outward push than the $^{235}UF_6$. The required centrifuge technology is quite sophisticated. Concern has sometimes been expressed that the technology may be used to further purify the uranium until bomb grade quality is reached.

In a nuclear reactor some of the neutrons may also react with the ^{238}U thereby producing plutonium. In the end a wide variety of radioactive elements results that remains in the reactor fuel after most of the ^{235}U has been used up. Some of these may be strongly radioactive but with short half-lives, and these are generally kept on the reactor site where they have to be well protected. Others may have lifetimes of thousands or hundreds of thousands of years which pose a major problem. It is now thought that these should be stored in geological repositories: deep tunnels in which they would be protected against water, intruders and other threats. In the USA a repository for radioactive waste has been planned for several decades now in Yucca mountain, but litigation has held up its actual implementation. The site in that mountain is well above the groundwater level several hundred meters below the surface. While there is general agreement that such a repository is needed, everyone wishes to place it in his neighbor's territory. Perhaps the most advanced are the Swedes. With a modest tax on nuclear electricity they wish to ensure that the present generation takes care of all expenses for permanently storing the nuclear waste. After all it is hardly reasonable for the present generation to enjoy the electricity but to leave to future generations the worry about what to do about the waste. By inviting everyone to come and visit the repository deep underground they have created a climate of trust and openness that has avoided controversy in the communities where it is to be located.

Conventional reserves of uranium are not very large with estimates going up to some 10 million tons. Since a 1-GW reactor needs some 150 tons of natural uranium per one-sixth of a year, the 63 TW annual requirement for the 100,000-year society would suffice for only six years. Less abundant ores (100 ppm) might provide several times more. However, it is thought that to obtain uranium from ores with less than 10 ppm of uranium would take as much energy as it would provide [24]. The world's oceans contain about 4,000 million tons of uranium and so might suffice for 2,400 years. While in Japan some experiments have been made in obtaining uranium from the sea, it would be a Herculean task to push the immense volume of water of all the oceans through a treatment plant in 2,400 years' time. So the conventional nuclear reactors are hardly a promising source of energy for the long-term future.

Since ^{238}U is so much more abundant than ^{235}U, much more energy could be generated if a way could be found to use it in a chain reaction. This is possible when it is transformed into plutonium ^{239}Pu. This transformation may be achieved with faster neutrons in a 'Fast Breeder Reactor', where more energetic

neutrons strike the ^{238}U. The net gain in energy output would be about a factor of 60. As a consequence, also more energy would be available to extract the uranium from more abundant poorer ores. In this way an important contribution to the energy needs of the 100,000-year world would become possible. But it would be a pact with the devil: plutonium is a powerful poison and the basis for making nuclear weapons. To have large quantities of it in the energy system would be an invitation to disaster.

In this respect thorium provides a much better option for a breeder reactor. ^{232}Th is a radioactive element with a half-life of 14,000 million years, and in the Earth's crust it is about five times more abundant than uranium. It is unsuitable for having a chain reaction, but in a fast breeder it can be transformed into ^{233}U which is suitable. The great advantage is that no plutonium is produced, though there is radioactive waste. A particularly attractive proposal has been made to generate some of the neutrons needed for the transformation externally. For example, a particle accelerator could produce energetic protons which then in a lead (Pb) target would generate the neutrons. The reactor then could be subcritical, which means that it produces slightly fewer neutrons than needed to maintain a chain reaction, the other neutrons resulting from the energetic protons. This represents an important safety feature [25]. If for some reason the reactor has to be stopped, all that needs be done is to switch off the particle accelerator. Thorium is extremely rare in the oceans because its oxides are insoluble, but it probably could be extracted from granites where it has an abundance of some 10–80 ppm. So the required amount of thorium to satisfy one-sixth of future energy needs could be obtained without too many problems. Nevertheless, because of the difficulties with waste disposal, it would be desirable to restrict its use to a more modest part of the total energy mix. It appears that much more research has been done on uranium breeders than on thorium breeders. However, a modest experimental reactor based on the usage of thorium has been running for several years in India, a country with large reserves of thorium ore [26].

7.1.7 Fusion energy

The same energy that powers the Sun could be a nearly inexhaustible source of energy on Earth. In the solar interior the temperature is high enough ($\sim$15 million $^{\circ}$C) for nuclear reactions to occur which fuse four hydrogen nuclei into one helium nucleus. This liberates about 180 TWh of energy per ton of hydrogen. Thus one-sixth of the annual energy requirement of our future society could be met with some 500 tons of hydrogen which may be obtained from 4,500 m^3 of water – an amount equal to the consumption of a village of a few hundred inhabitants! However, in practice things are not so simple, because of the difficulty of containing such a hot gas on Earth.

The first reaction in the production of helium fuses two hydrogen nuclei into one heavy hydrogen (D or deuterium) nucleus composed of one proton and one neutron. This reaction does not produce much energy. Because of some subtle nuclear effects it is so slow that even in the 4.5-billion-year-old Sun most of the hydrogen has not yet reacted. It is therefore better to begin the process directly

with deuterium. In fact, about 0.01% of ocean water is 'heavy water' where the hydrogen in H_2O is replaced by deuterium, yielding HDO or D_2O. There is thus an ample supply of heavy hydrogen. Different ways to achieve fusion are then open: D + D → helium or D + tritium → helium + neutron, which proceeds at a lower temperature and is much easier to realize. Then we need tritium, the third isotope of hydrogen with a nucleus made up of one proton and two neutrons. Since tritium is radioactive, with a half-life of only 12 years, it does not occur naturally on Earth, but can be made by striking lithium nuclei with neutrons. We then have as reactions:

$$\mathrm{D + T \rightarrow He + n}$$
$$\mathrm{n + Li \rightarrow He + T}$$

with the final result

$$\mathrm{D + Li \rightarrow 2He}$$

with a heat energy yield of around 8 GWyr per ton of lithium, corresponding to around 3 GWyr of electricity. It also has been suggested to utilize a reaction of deuterium with ^{3}He that is found on the surface of the Moon, where it has been deposited by the solar wind (see Chapter 9). Quite apart from the problems of extracting ^{3}He from the lunar soil, the total resource would yield only 2,000 TWyr of energy. With the same efficiency of conversion to electricity, this would correspond to no more than 75 years at 10 TW, one-sixth of the long-term energy need. Actually the energy gained might not even suffice to set up the whole infrastructure required.

The problem is to enclose the hot deuterium–tritium plasma. This cannot be done in a material vessel, since the gas would cool down immediately when the nuclei struck the wall. Instead, this may in principle be achieved by a 'magnetic bottle' in which the magnetic forces confine the charged particles. For the last 50 years plasma physicists have been struggling to realize such a magnetic bottle, but many instabilities have until now prevented full success. However, it is anticipated that effective plasma confinement will be demonstrated with the International Thermonuclear Experimental Reactor (ITER [27], see Box 7.3) which is being built at a total cost of some €12,000 million. ITER should be completed within a decade or so following which it will be used for various experiments to optimize it. If all goes well, then by 2025 construction of a prototype commercial reactor could begin and perhaps 15 years after that the first of a series of fusion reactors. Even if such a schedule is followed, fusion energy is unlikely to become significant to the world's energy balance before some time in the second half of the century.

To facilitate the plasma confinement, a fusion reactor would have to be large and typically produce 1–10 GW of electrical power. If one-sixth of the total energy requirement of 63 TW were to be met by fusion, some 1,000–10,000 fusion reactors would have to be built. Deuterium for these reactors would be easily obtained from the ocean, but the necessary 1,200 tons of lithium annually would exhaust currently known reserves within 12,000 years. A possible way to

Box 7.3 ITER, the International Thermonuclear Experimental Reactor

The nuclear fusion reactions can only occur when the nuclei approach each other with enough energy to overcome the repulsive effects of their electrical charges. So the hydrogen gas has to be hot, in the Sun about 15 million °C and in the low densities of a terrestrial reactor around 100 million °C. But how are we to contain such a hot gas? In the Sun's interior the solar gravity acting on the overlying layers does this. On Earth some kind of vessel is needed. A material vessel would not do, because it would cool the gas or the hot gas would destroy it.

Since the nuclei are electrically charged they can be deflected by a magnetic field. Such a field is generated by electrical currents and, by arranging these appropriately, field configurations can be obtained that contain the hot gas. In ITER the hot gas will be confined in a toroidal configuration of the kind first developed in the 1950s in the USSR, the 'tokamak'. Unfortunately, most magnetic configurations become unstable when the contained gas exerts some pressure. For the last half century slow progress has been made in studying these instabilities and in finding ways to suppress them.

The most promising reaction is that of deuterium (D) with tritium (T), two isotopes of hydrogen (see Figure 7.2). Tritium is radioactive with a 12-year half-life and so does not occur naturally on Earth. However, it can be made by letting neutrons (n) react with lithium (Li) nuclei. Conveniently, neutrons are produced by the D–T reaction. Since neutrons have no electrical charge, they are not deflected by the magnetic field and strike the wall of the containing vessel. If that vessel is made of lithium, the necessary tritium is produced according to the reaction $Li + n \rightarrow He + T + n$, with the second neutron having a lower energy than the first.

A fusion reactor has to be big in order to sufficiently suppress the losses due to magnetic instabilities. As a consequence, even an experimental facility is extremely expensive. ITER was proposed in 1985 as an international project by Mikhail Gorbachev – at the time General Secretary of the communist party of the USSR – to the American President Ronald Reagan during their summit meeting in Geneva which sealed the end of the cold war. In a way ITER is the successor to JET, the Joint European Torus, which came close to the break-even point between the energy input into magnetic fields and particle heating on the one hand, and the output from the nuclear reactions on the other. ITER should be the first reactor with a net energy output. The project has been joined by China, India, South Korea, Russia, the USA (which withdrew in 1999 and rejoined in 2003) and the European Union which, as the host, pays about half of the USD 12 billion cost. After prolonged controversy ITER will be located at Cadarache in southeast France. A related facility for materials research will be placed in Japan. ITER is optimized for experiments. It is to be followed by DEMO, a prototype for commercial power production. Fully commercial fusion reactors could hardly be completed by 2050, unless a much larger investment were made and greater financial risks accepted.

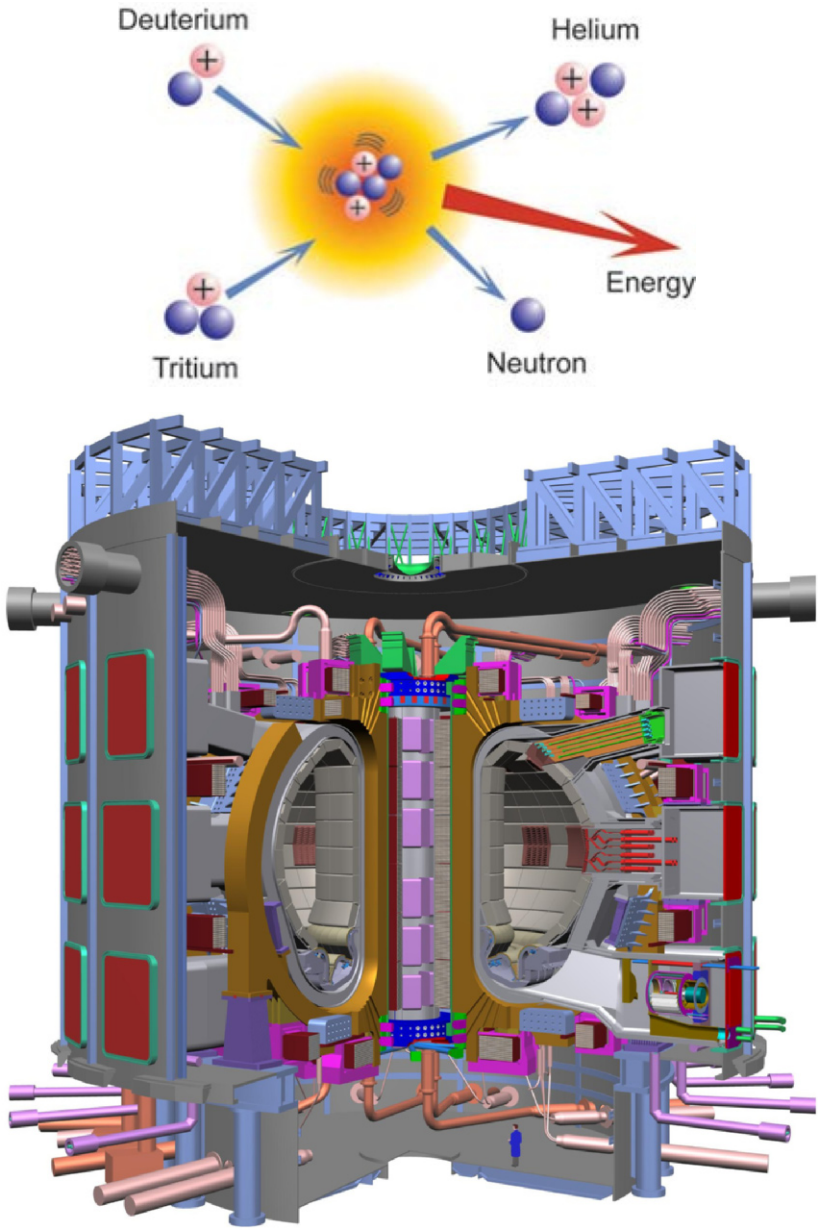


Figure 7.2 Fusion energy. Above, the D–T reaction with protons in rose and neutrons in blue. Below, a model of ITER. The high vacuum torus in the middle with its elongated cross-section will contain the 100 million °C D–T plasma which at any time will have a mass of no more than a few grams. The small figure at the bottom indicates the scale. (*Source*: ITER.)

fuel the fusion reactors would be to obtain the lithium from ocean water, where each cubic kilometer contains 170 tons; 0.05% of the oceanic lithium would suffice for 100,000 years. Moreover, there would be still much more in the Earth's crust.

It is important to note that a fusion reactor is intrinsically safe. In the 1,000-m^3 reactor vessel there is never more than a few grams of the reacting matter (Figure 7.2). Any disturbance in the vessel will cause this hot matter to reach the wall where it would immediately cool, terminating the fusion reactions. It is, of course, true that tritium is radioactive, but since there is so little of it a major accident is excluded. In addition, the walls of the reactor vessel will become somewhat radioactive by being struck by the neutrons. Choosing the right materials will minimize this, and the formation of the very-long-lived radioactive isotopes that create such a problem for fission energy would be prevented. So fusion may well have a bright future. Before we can be sure, however, it has to be proved that ITER will function as expected.

In conclusion, we can see many ways to power the 100,000-year society with perhaps 5% coming from geological sources, 10–20% from wind, 10–40% from the Sun, 10–50% from fusion, and 10–50% from thorium breeders. Both thorium and fusion reactors would produce the energy as heat that would then be transformed by turbines into electricity. In the future, more crowded, world the waste heat in this transformation might well be used for city heating and other purposes, thereby reducing the assumed 63 TW electricity requirement. If the photosynthetic efficiency of plants could be improved by factors of 10 or more, biofuels could make a contribution.

7.2. Energy for the present century

7.2.1 Fossil carbon fuels

Three carbon-based sources dominate the world's energy supplies: oil, gas and coal account for some 80% of current energy use. All three are the result of photosynthesis in the remote past. Photosynthesis is not a very efficient process with typical plants converting only 1% or less of the incident solar energy into chemical energy, but during the long periods of geological time enormous quantities have accumulated. Coal was the first to be extensively exploited, and because the deposits were widely distributed, most countries could mine their own coal. During the last century and a half oil and, more recently, gas have gained in importance because of greater convenience. But the distribution of oil and gas is much more concentrated in specific areas, which has led to political consequences. Oil and gas are easy to transport through pipelines or large tankers. Moving coal by ship and truck is more cumbersome. In addition, mining coal still has relatively high fatility rates from accidents and lung disease, but especially in developing countries without much indigenous oil, the use of coal is still increasing.

Coal is the result of the burial of land plants. Especially during the carboniferous epoch huge tropical swamps developed after robust plants had

evolved that could be buried before rotting away. In fact, this drew down much of the atmospheric CO_2 that contributed to the Permian ice ages. Oil and gas are mainly due to the decay of marine organisms. The result depends on the temperature history of the material after deposition on the ocean floor, subsequent tectonic movements and the migration through sediments. It is believed that typical petroleum reservoirs have been processed abiotically at temperatures of the order of 100–200°C, with natural gas resulting at the higher temperatures. In reservoirs at temperatures below 80°C, bacteria or archea may prosper leading to biodegradation of the oil. This would lead to 'non-conventional' oils, which are more viscous and harder to process. Included are the heavy oils of the Orinoco basin in Venezuela, oil shales and the tar sands of Alberta in Canada [28]. It is a curious accident of history that, although conventional oil is principally found in the deserts of the Middle East, more than 80% of the known non-conventional oil is situated in the western hemisphere.

Much of the natural gas comes from reservoirs in the Middle East and from Siberia, but there are much larger quantities on the bottom of the oceans in the form of methane hydrates – mixtures of frozen methane and water-ice – that may form under high-pressure, low-temperature conditions when methane gas seeps upwards [29]. Although this form of methane is believed to exceed in energy content all other carbon-based fuels, it is uncertain how much of it can be effectively recovered. The great attraction of natural gas is that it produces less than 60% of the CO_2 that results from coal for the same amount of heat energy. Also, fine dust and other pollutants are produced in the burning of coal. As a consequence, the newer power plants in the developed countries are frequently using natural gas.

The estimated overall global consumption during 2005 of oil, gas and coal is shown in Figure 7.3(a). It appears that oil is the dominant component because almost all transportation, some power generation, and home and other space heating depend on it.

Ever since oil became widely used, concerns about imminent shortage have been expressed. In *Limits to Growth*, it was foreseen that 'known global reserves' would be exhausted for oil in 2000 and for natural gas seven years later. However, this has not happened because new reserves have been discovered (a possibility already mentioned in the report) and the rate of growth of consumption has been lower than foreseen, in part by improved efficiency in energy use. In fact, present reserves are double what they were in 1970.

The availability of any resource depends on its cost. If oil is cheap, no one is going to make expensive efforts to find new supplies or to make the extraction more efficient; but if there is a risk of imminent scarcity, the price goes up and there is a motivation for increasing the supplies. The cost of energy world wide has been of the order of about 3% of total world Gross Domestic Product. For every 30 euros in the global economy only 1 euro was spent on energy. In that sense energy was cheap and resources that are harder to obtain are not exploited. In fact, such resources exist. Very recently oil prices have tripled, but it is uncertain if this will remain so.

The current oil supplies come from very favorable geological structures: over millions of years organic matter has been accumulating in shallow seas and subsequently been covered by salt deposits. Both were covered by sediments as time progressed. Over many millions of years the organic matter was transformed into oil. As the salt served as a seal, large reservoirs of oil were formed. Drilling through the salt, one may gain access to the contents of these reservoirs to tap the oil or gas. During past geological ages there was the super-continent Gondwana with, between its northern and southern parts, the equatorial shallow Thetys Sea (Figure 2.4) where conditions were particularly favorable for the formation of hydrocarbon reservoirs. Subsequently, northward continental motions moved these to what is now the Arabian Desert belt. Not surprisingly, these rather empty desolate regions with huge resources were the envy of the industrial powers with all the consequences one sees today. It is not at all clear that many such large reservoirs of easily exploitable oil remain to be discovered. Quantitative predictions are still uncertain and optimistic and pessimistic estimates alternate. One has only to look at the titles of articles in *Science*: 1998, 'The next oil crisis looms large – and perhaps close'; 2004, 'Oil: never cry woolf – why the petroleum age is far from over'.

Currently estimated reserves of energy are shown in Figure 7.3(d). These estimates include well-ascertained reserves but also much more uncertain estimates of resources of conventional oil and gas still to be discovered, which are based on previous experience and general geological understanding. In Figure 7.3(e) are shown the supplementary CO_2 concentrations in the atmosphere that would result from the burning of the resources. These are very uncertain as they depend on various aspects of the carbon cycle (Chapter 6) that have not yet been evaluated sufficiently precisely. These are the central values of different parameterizations; reality may be more favorable, but may also be worse. As explained in Chapter 6, current CO_2 concentrations in the atmosphere are around 380 ppmv, and it would be desirable to keep these below 450 ppmv in the future; so no more than 70 ppmv should be added. If this limit were to be exceeded, the result could be the ultimate melting of the Greenland Ice Sheet with the consequent raising of the sea level by some 7 meters. So even exploiting all the currently expected oil and gas would bring one into the danger zone and coal could exacerbate the problem.

As we noted before, non-conventional oil and gas resources may dwarf the conventional ones. In Figure 7.3(f) are indicated the estimates of the total of all of these. While it is not certain how much of the methane in hydrates may be recovered, it is seen that potentially the available energy could be increased by a factor of 18 with the increase in CO_2 perhaps being a factor of 10. However, such a CO_2 concentration would be far beyond the validity of our climate models.

The hydrate discoveries might solve our energy problems for a long time to come, if the technology for recovering the methane can be developed. They also give perhaps the clearest warning of the dangers of continuing the extraction of hydrocarbons to power our society. Geologists have discovered a remarkable event some 55 million years ago – the so-called Late Paleocene Thermal

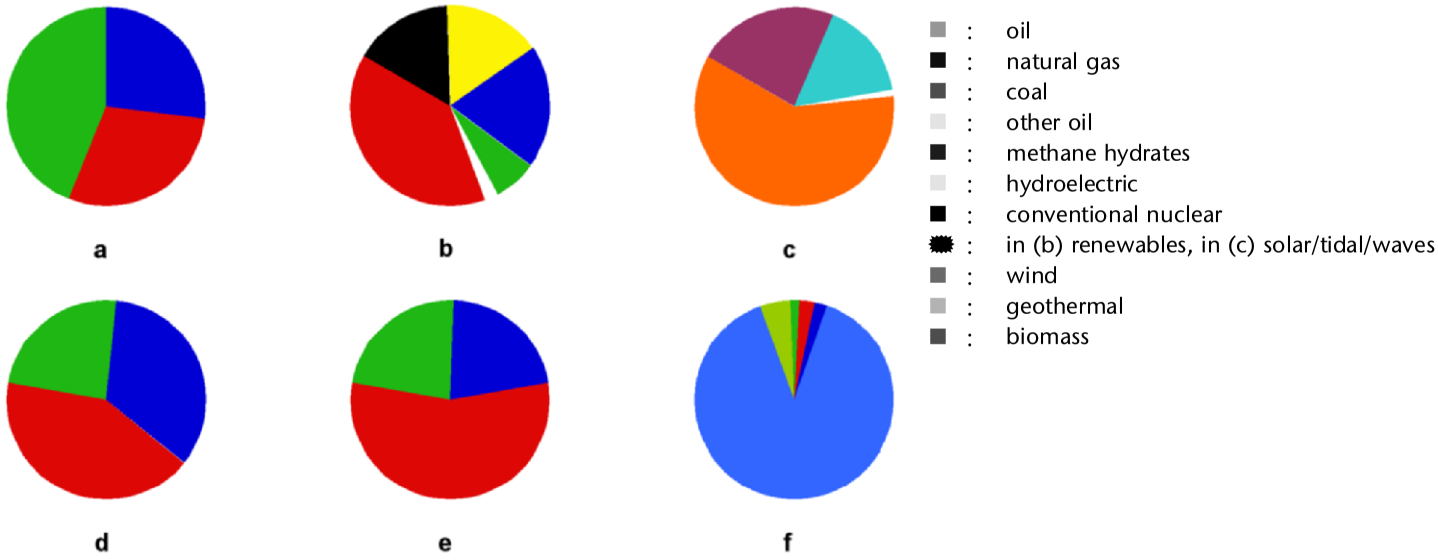


Figure 7.3 Current (~2005) energy production and supply. Each image indicates the distribution over various sources: (a) energy from hydrocarbons 370 EJ; (b) electricity generation 17,500 TWh from various sources (heat equivalent 63 EJ); (c) electricity generation 400 TWh from renewables (1.4 EJ); (d) the 45,000 EJ reserves of conventional hydrocarbons; (e) the resulting 220 ppmv increase of CO_2 concentration from burning all of (d), assuming an average climate model (the result has much uncertainty); (f) speculative ultimate energy availability from non-conventional hydrocarbons. It is not at all obvious what part of the 750,000 EJ can actually be exploited. (Data for (a)–(d) from OECD/IEA, USGS, CEA.)

Maximum. The sediments deposited at that time show that quite suddenly (no more than a thousand years) the ratio of the two isotopes of carbon ^{12}C and ^{13}C changed as if a large quantity of ^{13}C-poor carbon had been injected into the atmosphere. The most plausible explanation is that volcanic magma penetrated into the layers of methane hydrate (or possibly of coal) and that the resulting heating released the gas. Some 10^{15} m^3 must have entered the atmosphere where it would have been oxidized to CO_2 (see Box 6.2 on page 203).

At about the same time that the methane was released the temperature shot up by 4–8°C. The resolution of the geological record is insufficient to see how fast the temperature increase was, but it happened in less than a thousand years. About 100,000 years after the event the $^{12}C/^{13}C$ anomaly had been ended, presumably by the absorption of the atmospheric CO_2 into the oceans, and the temperature came down again. It is sobering to think that the methane injection into the atmosphere was only 5% of the total reservoir believed to exist today. However, the estimates of that reservoir are very speculative. The resulting CO_2 injection is about equal to the cumulative total expected to have been injected some 40 years from now by our consumption of oil, gas and coal.

This drives home the essential point. Our present-day energy problem is not that there are insufficient resources of oil, gas and coal, but that if we continue using them, the climatic effects may become unmanageable. Of course, should effective ways be found to fully and reliably sequester the CO_2 and thereby stop the further pollution of the atmosphere, the situation might change [30]. Such storage should be very perfect. If, say, 1% per year escaped into the atmosphere, the resulting global warming would only be delayed by some 50 years or so.

7.2.2 Electricity and renewables

Electrical energy is produced from combustion of coal, gas and oil, from nuclear heat sources, from hydroelectric plants and from a variety of other renewables, which at the present time account for only 2% (Figure 7.3 (b) and (c)). Both nuclear and hydroelectricity produced about 16% of the total 17,500 TWh (2002) of electrical energy. Since the conversion of heat energy to electrical energy typically has an efficiency of no more than 30–40%, electricity generation is a significant contributor to the CO_2 problem.

7.2.3 From now to then

In the preceding sections we have seen that the present sources of energy will not be available for the long-term future. So how and when are we going to make the change-over? Of course, the matter has gained urgency, because of the effects of present-day energy sources on the climate. At the same time, it has been frequently pointed out that it is difficult to make a rapid switch to renewables, because of the inertia of the energy infrastructure: aircraft live 20 years and power plants half a century. Some have concluded from this that one may well wait a bit longer. More reasonably one could argue that it is important to make new power plants suitable for the future, even if this were to cost some more in the short term. If the full effects of climate change are included in the

calculation, it is not even obvious that there are extra costs. Also, it seems clear that in the not too distant future hydrocarbons are going to be more expensive. In addition, future intensive use of hydrocarbons is only acceptable if CO_2 is safely stored away, which entails further costs. Hence, a cost comparison based on present-day prices may be far off over the lifetime of a power plant. We have seen that, in the long-term, nuclear energy is hardly a viable option. If most of our energy would be generated from natural uranium the supply would not last long, while the danger of the plutonium produced in the breeders seems prohibitive. Fusion has a great potential, but the construction of the first fusion reactor is at least several decades into the future. With limited possibilities for hydropower and geothermal energy, wind, solar and biomass will have to be the main additions. Technologically wind turbines seem well established and economically not too far from competitive, while solar cells will need further development, since, for the moment, their electricity is 10 times more expensive than the eolian. The large-scale use of biomass will need the further development of more productive plants.

What would be involved in obtaining a significant impact from wind power? World production of electricity amounted in 2005 to around 18,000 TWh, increasing by some 3% per year. Let us modestly assume that we would decide to cover half of the increase by wind energy. A large wind turbine in a suitable location generates some 5 GWh per year. Half of the annual increase in electricity consumption amounts to 270 TWh. As a consequence, we would have to build 54,000 wind turbines each year, which would require an area of 8,000 km^2. Wind power in the world in 2005 generated 94 billion kWh. So, if we were really serious about wind energy to take up half of the increase in electricity consumption, we would have to install each year three times as much as the 2005 cumulative world total.

Current cost of each turbine would be of the order of USD 1.5 million. With the mass production of turbines, cost could still come down a little to say, USD 1 million, and so the annual cost would be USD 54 billion. This seems to be very high, but of course if we build conventional or nuclear reactors the cost will not be much less. Moreover, a tax of 0.3 cent per kWh produced by fossil fuel would cover the requirement. After 25 years one still would have only some 20% of all electricity from wind, and so the problem of intermittency would not, at that time, be too serious. While one may argue about the precise numbers, it is clear that unless present-day efforts towards renewables are increased by more than an order of magnitude, no significant effect on global warming will be achieved.

In parallel with the implementation of wind energy, an enhanced program of research on photovoltaics and biomass would be required in the expectation that a larger scale implementation of these methods would become possible, which could then begin to reduce the other half of the increase in electricity production, currently foreseen to be provided by fossil fuels. The issue is not to close down functioning fossil fuel plants, but not to build more new ones than minimally required. In this way fossil fuel generated electricity might be phased

out by the end of the century and the implementation of ambitious uranium/plutonium-based nuclear programs avoided.

7.3 Elements and minerals

More than 90 different chemical elements are known to occur in nature. While the simplest, hydrogen, is made of atoms composed of one electron orbiting a nucleus, the most complex, uranium, has 92 electrons and a very heavy nucleus. The chemical properties depend mainly on the number and arrangement of the electrons around the nucleus, which determine the compounds or minerals that the elements can make. Metals are elements in which the electrons have much collective freedom of movement and therefore have high electrical and heat conduction. Some 22 elements are directly necessary to human life [31]. Almost all elements have important technological uses. Some elements are very abundant on Earth: for example, oxygen is an important constituent of the atmosphere and the oceans, while with silicon it accounts for three-quarters of the Earth's crust. A very rare element, like gold, contributes only a milligram per ton (1,000 kg) of crustal rock.

7.3.1 Abundances and formation of the elements

Much of the matter in the Universe consists of only two elements: hydrogen and helium. All other elements together contribute about 1.5% by mass. The same applies to the material of which the Sun is made (Figure 7.4). Eight elements make up almost all of the 1.5%, with all the rest contributing no more than 0.015% [32]. When the Universe was still very young, it was very hot; and even the protons (hydrogen nuclei) had not yet appeared. But as the Universe expanded, the temperature fell, the protons formed and, subsequently, the helium nuclei and minuscule quantities of other elements. Because of the rapid temperature decrease and the relatively low density, no further nuclear reactions occurred. Much later, stars began to form and high temperatures were reached in their interiors at much higher densities than in the early Universe. This made it possible to transform some of the hydrogen into additional helium. More importantly, it allowed heavier elements to be synthesized.

The Sun and most stars obtain the energy they radiate from their surface by converting hydrogen into helium in their deep interiors. At some moment all the hydrogen there will have been used up, and what will happen then? The star will continue to radiate, but no energy will be supplied, so we could expect the star to cool down as it radiates away its heat energy.

If the star were to cool, the pressure of the gas in its interior would diminish, but this is exactly the pressure that prevents the star from collapsing under its own gravity. Therefore, as the star cools, it will contract, compressing the gas in its interior and this compression will heat up the gas. This will lead to the paradoxical result that as the star continues to radiate without a supply of nuclear energy, it will get hotter rather than cooler. As the stellar interior heats

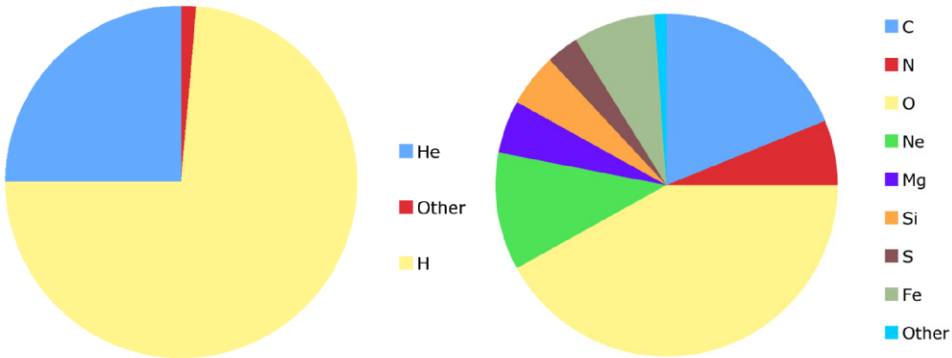


Figure 7.4 The abundances of the elements in the Sun. On the left the red segment corresponds to all elements other than hydrogen and helium. The relative abundances of those are shown on the right.

up, the temperature may become so high that more complex nuclear reactions become possible: three helium nuclei may fuse to make a carbon nucleus, and with the addition of one more, an oxygen nucleus. Also, as the carbon may be mixed into a part of the star that still contains some hydrogen, nitrogen could be formed. Gradually in this way several of the elements we know on Earth could be synthesized, but as they would still be in the deep interior of the star, how could they become accessible?

Observations and modeling show that stars may become rather unstable during their evolution. This progression may lead not only to the mixing of matter in the interior and at the surface, but also to the ejection of shells of matter. As this ejected gas mixes with the interstellar gas, the latter is enriched in the elements that have been synthesized in the stellar interior. As new stars form from this gas, they will already possess a certain content of such elements and may later eject even more. In the course of several generations of stars the composition of our Sun was determined. So during its formation the elements needed to form planets had also become available.

This, however, is not the complete story. Energy is liberated by the fusion of nuclei, but only up to iron. Heavier elements require energy for their synthesis. Suppose we now have a star with an iron core. It will continue to radiate, but as it cannot generate the required energy by nuclear reactions, it will therefore continue to contract. At that stage the stellar interior loses its stability and collapses to form a neutron star, or in some cases a black hole. The collapse releases an enormous amount of gravitational energy, enough to explosively heat and eject the overlying stellar envelope. During this explosion, which lasts only a few seconds, a wide variety of nuclear reactions take place and many elements are synthesized and ejected. In lower mass stars such explosive events may also occur even before an iron core is formed. The energy deposited in the envelope also leads to a very high luminosity up to several thousand

million times the luminosity of the Sun: and a supernova appears. Supernovae are rare, as most stars end their lives differently by losing matter more slowly.

The course of events we have sketched for the synthesis of heavy elements has been confirmed by the 1987 supernova in the Large Magellanic Cloud – the most extensively studied supernova in Earth's history. Before the supernova explosion, a rather faint star had been observed, but it was not particularly noteworthy. Nothing indicated that in its interior the final evolutionary phases were taking place, until, on the morning of 23 February, a burst of neutrinos was detected in Japan and in the USA which signaled the core collapse. Some hours later the stellar luminosity increased rapidly to reach a maximum of more than 100 million solar luminosities, followed by a slow decline. Several clear indications of element synthesis were subsequently found.

From the properties of the nuclei of the elements involved in the reactions it follows that nickel will be produced in substantial abundance. Nickel (nucleus with 28 protons), found on Earth or in meteorites, is mainly composed of two isotopes with, respectively, 30 and 32 neutrons. However, the principal isotope produced in the explosive supernova reactions has only 28 neutrons. It is unstable and decays with a 6-day half-life to a cobalt isotope (27 protons, 29 neutrons) which is also unstable and decays with a 77-day half-life to the most common stable iron isotope (26 protons, 30 neutrons). During the first few weeks, the heat generated in the explosion still lingers, and all the radioactive nickel decays. Subsequently, it might have been thought that the supernova cools and fades rapidly but, instead, it is seen that the brilliance of the supernova declines slowly – being halved every 77 days. The explanation is simple: the decay of the cobalt (resulting from the nickel decay) heats the supernova envelope and since the quantity of cobalt is halved every 77 days, so is the energy input into, and the radiation from, the envelope. From the luminosity of the supernova of 1987 it follows that an amount of 0.07 solar mass of nickel has been generated in the supernova which, after some months as cobalt, decayed into stable iron. This scenario has received a striking confirmation from spectroscopic observations of the supernova. After a few months, when the outer envelope became transparent, a spectral line due to cobalt appeared and subsequently weakened due to the radioactive decay.

Iron and silicon, with oxygen the main elements in the construction of the Earth, have been largely synthesized in supernova explosions. The same is the case for 14 of the 22 elements needed for human life. Others have been synthesized during the earlier, calmer phases of stellar evolution. We are truly children of the stars.

A detailed analysis of the nuclear reactions that may occur, and of the physical conditions that may be encountered in stars, has shown that most elements and isotopes found in nature can be synthesized in stars. For a handful of isotopes, however, this is not the case. Some of these have been formed during the early hot phases of the Universe. The most important are deuterium (heavy hydrogen), and most of the helium and lithium isotopes.

More than 99% of the mass of the Solar System resides in the Sun, and so its

composition should be representative of the average composition of the Solar System. From an analysis of the solar light, the abundances of the elements in the Sun may be inferred (Figure 7.4). Two elements, hydrogen and helium, account for more than 98% of the solar matter; both are rare on Earth. This should not surprise us. Both elements are gaseous and even if they had been present when the Earth formed, the Earth's gravity would have been insufficient to retain them, except for the hydrogen bound in heavier molecules like water, H_2O (see Chapter 2).

When the Solar System formed most of the matter went into the Sun. But some of it formed a disk, extending sufficiently far from the Sun to be relatively cool. Here solid material could condense. Gradually these solids coalesced into larger and larger bodies (Chapter 2). The largest of these had the strongest gravity and so attracted their neighbors most effectively, growing very rapidly. These became the planets, while the surviving small bodies formed the asteroids. Some of these collided with each other and broke up into fragments. From time to time, such fragments fall to Earth as meteorites.

Some meteorites are mainly made of iron and nickel. They should be the fragments of modestly sized bodies, originally hot enough for these elements to melt and separate out from the rock as the heavier parts sunk to the middle. Some are stony – like the Earth's crust – and are fragments of the crusts of these 'planetesimals'. Finally, there is a class of meteorites – called 'chondrites' – that seem to be homogeneous. These come from smaller planetesimals that never have been hot enough to become chemically differentiated, because they were too small. If we exclude the gaseous elements, we find that the composition of the chondrites is the same as that of the Sun, which confirms that their composition is representative of the original composition of the Solar System insofar as the non-volatile elements are concerned. Several very rare elements cannot be detected in the solar spectrum, but can be measured in the chondrites, and in this way their abundance in Solar System material can be ascertained.

7.3.2 The composition of the Earth

The overall composition of the Earth should be expected to be the same as that of the Solar System material from which it formed, except for the loss of the volatiles. So the whole Earth composition should be the same as that of the chondrites. Since the Earth had sufficient mass to become very hot during its formation, we then could expect that the abundant heavy elements, iron and nickel, have melted and dripped down under the influence of the Earth's gravity to form the core, leaving the lighter silicates to be the dominant constituent of the crust (Chapter 2). Due to their chemical properties some elements have a notable affinity for each other and therefore tend to segregate the same way. Thus, much of the 'iron loving' (siderophile) elements went with the iron into the core; examples are sulfur, nickel, gold and platinum which, as a result, are much depleted in the crust. The 'stone loving' (lithophile) elements with the opposite chemical characteristics – such as silicon, potassium or uranium – concentrated in the crust.

Chemical differentiation also occurs on much smaller scales. For example, if oceanic crust is dragged down to greater depth owing to the motion of continental plates, hydrothermal fluids (water with sulfur and other substances) will be slowly pushed up. During such a process the gold and other rare elements may be leached out of the rocks and dissolved into the fluids. When the fluids rise through cracks in the crust, temperature and pressure diminish, the gold is no longer soluble and is deposited at a location where the temperature has a specific value. So, in a limited area and at a specific depth, a gold deposit is created [33]. Later, when that area is lifted to greater heights and the overlying rock erodes away, the gold may appear at the surface or at a modest depth. When later the rock weathers, the gold nuggets contained therein will be unaffected and may be transported by rivers and streams far from where they came to the surface. The fabulous gold deposits along the Klondike river in the Canadian Yukon had such an origin, but gold is just one example. Many other rare elements may be concentrated by a variety of processes involving liquids from deeper down with different compositions. The result of these processes is that rich mineral deposits of economically important elements are very inhomogeneously distributed. For example, most of the world's reserves of chrome, cobalt and platinum are found in southern Africa, while much of the world's tin is found in South-East Asia. This has led to political tensions or wars as rapacious neighbors and others coveted valuable and unique deposits.

7.3.3 Mineral resources

Minerals have assumed an ever-increasing importance in human history as the technology was mastered to mine and extract valuable elements. The ancients obtained from their mines copper, iron, lead, gold, silver, tin, mercury and other elements. They also made bronze by combining copper found in Cyprus with tin mined in Cornwall and elsewhere – a technology that was developed independently in several places in the world. Today all non-volatile elements in the Earth's crust are being mined, with production ranging annually from iron at nearly 1,000 million tons to scandium at 50 kilograms. Some elements are essential to contemporary society: without iron and aluminum it would be impossible to construct our buildings, trains and planes, while several other elements are needed to make different kinds of steel. More recently several rare elements are in much demand by the electronics industry. Other elements are needed to make fertilizers for our agriculture, especially phosphor, potassium and nitrogen. But there are others that we could do without, although it might be inconvenient. Certainly a world without scandium should not give us sleepless nights! And, of course, there are the 22 elements that are needed to maintain human life, although most of them are required in small quantities.

In the early years of human mining activities very rich ores were exploited. In many cases such elements as gold, silver and copper could be found in pure form. Somewhat later rich mines with some 50% iron content or perhaps 5% copper were opened up. Most of these rich ores are now gone and the prospects of discovering significant additional amounts are not very bright. So the miners

and metallurgists have learned to find and process poorer ones, with sometimes no more than 0.3% of metal content in the case of copper. This has also increased the environmental damage due to mining activities. At 0.3% copper content, there are 333 tons of waste rock per ton of copper, plus eventually a large amount of overlying rock in the near surface mining process. With sulfuric acid being used to leach the copper from its ore and mercury to extract the gold, the wider environment of mines is frequently polluted and the long-term effects may be felt far downstream along rivers. The exploitation of poorer resources has required increasing amounts of energy for the milling and transportation of the ore.

For some time now concern has been expressed that we may 'run out' of essential elements. The resources in the Earth's crust are finite and on timescales of millennia are not renewable. This was stressed in *Limits to Growth* [34]. In the part that dealt with resources it was concluded that '*given present resource consumption rates and the projected increase in these rates, the great majority of the currently important non-renewable resources will be extremely costly 100 years from now*'. In fact, the reserves of 16 important metals were projected to be exhausted within a century and 10 of these by the year 2005. Fortunately, this has not happened because in the meantime new reserves have been discovered and technical developments have allowed lower grade ores to be exploited. In that sense the concept of a fixed amount of reserves is inappropriate. As technology improves, ores that could not be exploited may at a later stage become part of the reserves. This in no way invalidates the warnings of *Limits to Growth* as was joyfully proclaimed by the growth lobby. It only means that we have somewhat more time than was expected, because consumption in many cases increased less than foreseen and also because some new resources were found. The essential warning about the finiteness of the Earth's resources remains as valid as ever.

The fear of shortages caused much concern and led to the 'cobalt crisis' when instability in the Katanga province of Zaire suggested that the production might be much reduced. Even though no real shortages developed, it caused some industrial countries to organize strategic stockpiles of critical elements. In that climate of uncertainty, the German chancellor H. Schmidt in 1978 ordered a study to be made which concluded that if five critical elements became insufficiently available, 12 million German workers would lose their jobs. Not surprisingly, the study was immediately classified! [35]. In the meantime, those elements are still amply available and many of the stockpiles have been re-sold. It cannot be excluded that such fears will resurface again in the later parts of the present century.

Some 14 years after *Limits to Growth*, there appeared an article 'Infinite resources: the ultimate strategy' which, as the title indicates, came to a much more favorable conclusion [36]. It noted that seven elements could certainly be obtained in 'infinite' quantities from sea water and that for four others this probably would be the case. For six gases in the atmosphere (N, O, Ar, Ne, Kr, X) the same would apply. Also five elements in the Earth's crust (silicon from sandstone, calcium from limestone, aluminum (+ gallium) from clay and sulfur)

would be in 'infinite' supply. The same would apply to chromium, cobalt and the six elements of the platinum group if the technology to extract these from 'ultramafic' (basaltic) rocks could be successfully developed. For good measure iron might have been added to that list, since it is quite abundant (4% of the Earth's crust); so it is difficult to believe that it will ever become irrecoverable. Also the manganese nodules on the ocean floor and hydrothermal deposits there contain important quantities of some elements [37] even though their exploitation may cause environmental problems [38]. Making some extrapolations involving population growth and per capita consumption, the authors found that, of a total of 79 elements, only 15 would be exhausted by the year 2100.

7.3.4 The present outlook

The most complete data set concerning the worldwide production and availability of metals and other minerals is provided on an annual basis by the US Geological Survey (USGS) and made available nowadays on the Internet [39]. It is important to understand the terminology used: 'Reserves' denote the total quantity of a metal or mineral that may be mined with current technology at more or less current cost. The 'Reserve Base' is larger and includes currently marginally economical and sub-economical deposits. Typically it is twice as large as the 'Reserves'. 'Resources' may include deposits that have not yet been discovered but that, on geological grounds, are believed to exist, or are currently not economical, but have a reasonable prospect of becoming exploitable with foreseeable technological developments. As an example, for zinc the reserves are presented (in 2005) as 220 Mt (million tons), the reserve base as 460 Mt and the total resources as 1,900 Mt. For comparison, in 1970 the known global reserves of zinc were listed as 123 Mt and the annual production as about 5 Mt per year which, in the meantime, has increased to 10 Mt per year. So, in the intervening period some 200 Mt have been mined. The conclusion is that new discoveries may have added some 300 Mt to the reserves. For many other metals the situation is similar. In a way it is evident that gradually resources become 'reserves'. Part of the resources had been inferred but not yet discovered; technology has been further developed; and 'economical exploitability' is an evolving concept. Undoubtedly these factors may also increase the 'resources' further, although if, in lower grade ores, technological barriers to the recovery of an element are met, this need not continue to be the case.

If we take the estimates of the USGS in 2005 for the reserve base and assume that mine production would continue to increase at the same rate as over the last 15 years, we would find that some nine elements, including copper and zinc would be exhausted by 2050. However, for four of these the resource estimates of the USGS are substantially higher than the reserve base, and, for two of them, higher resource estimates have been made in the literature, leaving only gold, silver, antimony and indium potentially in short supply by 2050. All four elements are important in the electronics industry.

The case of gold is peculiar. Some 85% of the gold mined during human

history is still around in central banks, jewelry, etc. In total this represents 127,000 tons, more than the current reserve base in the ground. Gold is used in the electronics industry, but if there were a serious shortage, then some of the 127,000 tons could certainly be used, corresponding to another 50 years of production at current rates. Perhaps the situation is somewhat similar for silver and tin, with table silver and tin vessels still widespread.

Around the year 2100 most of the present resources of copper, zinc, the platinum group, tin, fluor, tantalum and thorium would also have been exhausted. Copper is a particularly important element in electrical engineering, agriculture and many other applications. Present 'reserves' will last only 21 years at the current 3.5% annual increase in consumption, and the 'reserve base' will last only 33 years. Recently the USGS has quadrupled its overall resource estimates to 4,000 Mt which could suffice for some 70 years if the increasing consumption flattens off at about five times present. There are several purposes for which copper may be replaced by zinc; roofing and gutters are examples. In fact, if one looks at the construction of houses one sees how, in the course of time, copper was used when it was cheap and replaced by zinc when it was too expensive. However, this does not solve the problem: copper and zinc could *both* be in short supply at the end of the century.

The platinum group elements (platinum, palladium, rhodium, ruthenium, iridium and osmium) are in high demand in catalytic converters. They are also important in the electronics industry, as is tantalum. Another element that may no longer be available by 2100 is helium, which is found in some sources of natural gas, especially in the USA. It has a number of applications (including in balloons) of which the most important is in reaching the very low temperatures required for superconductors. Superconducting cables transmit electricity without losses and could, at least in principle, allow wind or solar energy to be transported efficiently to the end users.

The conclusion from the foregoing discussion seems clear. While there may be a few elements in short supply, overall there should, by 2100, be no major shortage of the metals and other minerals on which our industrial society depends. Prices will continue to fluctuate as a result of real or perceived temporary shortages due to lack of foresight, political events and speculation. As the cost of energy increases, a steady upward pressure on costs will become noticeable. Of course, for particular elements for which new unanticipated uses are found, shortages and major price increases may occur; an example is indium which, until recently, had limited uses, but owing to its role in flat screen displays, saw its consumption increase 10-fold and its cost even more over a period of just a few years [40].

7.3.5 Mineral resources for 100,000 years

Though most minerals should be sufficiently available through the year 2100, in the two or three centuries thereafter the supply will become more problematic. We shall have to extract the elements from ever leaner ores, the composition of which in the long run should approach crustal material (Figures 7.5 (a) and (b)).

In addition, about a dozen elements may be obtained from sea water (Figure 7.5 (c)). The other elements occur in minerals that are not soluble in water and only the materials from the Earth's crust will do, unless unreasonable amounts of water are used.

The first question to settle is: How much will be needed of each element? When we discussed long-term energy consumption of the world, we found it to be seven times larger than in 2002 owing to increases in the world's population and the fact that the less-developed countries were reaching the same consumption level as the developed world. Very approximately, we could expect mineral and energy consumption to follow parallel trends – at least when averaged over some decades, and we shall therefore assume that mineral consumption in the 100,000-year world will also be seven times larger than in 2002. Again the year 2002 has been chosen to avoid the current turbulence on the natural resource markets. In fact, these are in large measure due to a rapidly increasing consumption in the developing countries. Thus to avoid double counting, we also take 2002 as our base year.

Technologically there should not be too many obstacles in extracting the 13 elements present in sea water with abundances [41, 42] of the order of 0.06 ppm or more. In the next chapter we shall see that, in the future, some 1,000–2,000 km^3 should be desalinated to obtain fresh water. The remaining brine will contain all that is required of 10–11 elements and a significant contribution to two others, calcium and fluorine.

There are huge layers of calcium carbonates and sulfates that were deposited in the seas of the past, and so there should be no shortage of calcium; most fluorine should come from rocks on land. However, neither the main construction materials, iron and aluminum, nor the heavier elements on which the electrical and electronics industries are based, can be obtained from the oceans.

The oceans contain more than 1 billion km^3 of water and so during 100,000 years at 1,000 km^3 per year, less than 10% would have passed through the desalination process. Thus we do not have to worry about it being a finite resource. Moreover, we shall have to return most of the desalination products since we obtain too much: for example, in 40 km^3 of sea water there is all the salt (NaCl) that will be needed each year. Also, through the rivers much of the elements we have 'consumed' will be returned to the sea. As an example, each year 400,000 tons of iodine are deposited on land by the spray from ocean waves and are returned to the sea by rivers [41]. The long-term iodine supply from desalination is only a third of the flux through the natural cycle.

We next turn to the elements obtainable from crustal rocks [41, 42]. Iron is certainly the most essential element for the industrial society. It accounts for 4% of the average crustal material, and in basalt it is about twice as abundant. In 2002 about 600 Mt of elemental iron were produced by mining world wide. So, with our assumption of a seven-fold increase in the future mineral consumption, we should then produce about 4,000 Mt of iron annually. With an iron abundance of 4%, some 100,000 Mt of average crustal rock would have to be

Figure 7.5 The abundance of the elements in the Earth's crust (a) and in the oceans (c). The 'other' segment in (a) has been expanded in (b).

processed annually, or half as much, 50,000 Mt, at the 8% abundance level in basalt. The latter figure corresponds to a volume of somewhat less than 20 km^3 of rock. In 100,000 years the total would become 2 million km^3 or 15% of the Earth's land surface to a depth of 1 km – a lot of rock, but theoretically not impossible. Great care would be needed to deal with the huge amounts of dust.

Having come this far and having secured the iron supply, we can ask what else we can extract out of the 50,000 Mt of rock. Aluminum could be produced in the same amount as iron, but its current need is a factor of 25 lower, which could produce a waste disposal rather than a supply problem. If we then look at the other elements in those 50,000 Mt of rock, we find a very mixed picture with also some important shortages (Table 7.2) as follows:

(1) Fortunately the principal elements needed in fertilizers for agriculture would be sufficient: potassium, phosphor and nitrogen, the latter obtained from the atmosphere. Also the small amounts of agriculture trace elements would not pose much of a problem. In fact, this is not really surprising since life could hardly have been based on substances that are very rare in the Earth's crust.
(2) While our sample has been chosen so that iron would be just sufficient, the element as such has limited uses. The industrial society utilizes a wide variety of steels – alloys of iron and other elements that give the required properties such as hardness, strength, resistance to corrosion, etc. Of the main elements needed for such alloys manganese, cobalt, chrome, vanadium and nickel would be sufficient, but tungsten and tin would not. Also, the quantities of zinc required for galvanizing against rust would be inadequate.
(3) Copper is the basic element for the electrical industry, but it would fall short by a factor of more than 10.
(4) The electronics industry utilizes a variety of heavy elements: gold, silver, antimony and others would be in short supply. If solar cells were to become an important source of energy, the requirements for such elements as indium would soon be unavailable and substitutes would be needed.
(5) Several heavy elements are used to catalyze chemical reactions in industry and in pollution control. The platinum group elements, rhenium and others, would be in short supply.
(6) Curiously the 'rare earth' elements would be amply available. This group of 15 elements with nearly equal chemical characteristics, which includes lanthanum and cerium, is actually rather abundant in the Earth's crust. They are used in catalyzators, glass manufacture and other applications.

The technologies needed to extract the elements with very low concentrations have not yet been developed. In fact, in some cases the abundances are very low compared to the minimum ore grade presently considered exploitable by a factor of more than 1,000. We could imagine that, to obtain the rare elements, one would at least have to melt or to vaporize the 50,000 Mt of rock; the latter would require some 15 GJ of energy per ton or 750 EJ in total. It is also very uncertain

Table 7.2 Elements that would be scarce in 50,000 Mt of basaltic crustal rock + 1,000 km^3 sea water. With 80% efficient recycling. only one-fifth as much rock and ocean water would have to be processed

Availability less than 10%:	Copper, zinc, molybdenum, gold, silver, tin, lead, bismuth, cadmium
Availability 10–50%:	Chrome, tungsten, platinum group, mercury, arsenic, rhenium, selenium, tellurium, uranium

whether this would be adequate, since the technology has not yet been developed, and the future world's energy consumption may even have greatly increased. Unless one can find effective extraction technologies it would be difficult even to process as much as that amount of rock.

There is one fundamental difference between energy and mineral use. Most energy ultimately winds up in the atmosphere and some in the rivers where it very slightly increases the temperature. So, as a result of its use, it is lost to further human consumption. While the metals and other minerals may accumulate in our garbage heaps, they are not lost and, as a consequence, may be recycled. In fact, recycling plays an important role today. Almost half of the aluminum is recycled at an energy cost of only 5% of that needed to extract it from ore. In the case of copper, some 15% is recycled. At present, recycling is an afterthought in the industrial world. A fundamental change in industrial design philosophy in which recyclability would be a priority could allow much higher percentages. If 80% could be reached, the 50,000 Mt of processed rock could be reduced to 10,000 Mt. The area needed would become 3% of the Earth's land instead of 15%, and the energy needed would become more plausible. With 90% recycling, a further factor of 2 would be gained. Whether this would be possible remains to be seen. However, from Table 7.2 it is clear that a number of elements will be in short supply even on favorable assumptions, and substitutes may have to be found.

So far we have assumed that further use of minerals is only affected by population growth and development. Of course, other factors may very much affect consumption. One example is the case of mercury: because of its ill effects in the environment, its usage has much diminished. At the time of *Limits to Growth* consumption was around 9,000 tons annually and increasing at 2.6% per year; in 2005 mine production had declined to 1,100 tons. On the other hand, as mentioned earlier, annual indium use increased 10-fold because of its use in flat screen television sets. So the simple assumption underlying Table 7.2, that future consumption increases by the same factor, is probably not really valid, although it indicates qualitatively the elements for which shortages are most likely to occur.

Of particular importance are the elements needed for energy generation. Fortunately thorium is abundant in granites, while recently the lithium abundance in the crust has been revised upward by a factor of 2 [43]. There would thus seem to be an ample supply of these elements for nuclear and fusion reactors.

7.3.6 From now to then

For quite some time the main measure that needs be taken will be to maximize recycling. To some extent this will happen automatically. As both energy and the richer ores become more expensive, the economics will favor the recycling option. Just as an example in the case of copper and aluminum, the energy required to process recycled metal is only some 10% of that of mined ore. As the latter becomes lower grade, the difference will become even larger. Nevertheless, a certain amount of governmental encouragement or regulation will be useful in stretching out the time before the more drastic measures discussed in the previous section become necessary. Moreover, any reduction in the amount of ore that is to be processed is beneficial, because the mining and refining of metals is a highly polluting activity and is very demanding in energy.

7.4 Conclusion

From the preceding discussion it appears that there is no insurmountable problem to having an adequate supply of renewable energy. Solar energy could produce all that is needed. It requires an energy storage medium, which could be hydrogen or some other vessel yet to be developed. Fusion energy is likely to be an almost inexhaustible source; as soon as ITER is shown to function as expected, this would become an important, if not dominant, component of the energy system.

The situation with regard to minerals is more ambiguous. Some elements will have to be replaced by others, and while this may be inconvenient, it would not bring the long-term society to a standstill. Much technological development will be needed to the extraction of elements from crustal rocks and to efficiently recycle all materials. The better part of a century is still available for such development.

7.5 Notes and references

[1] Ausubel, J.H., 1996, 'The liberation of the environment', *Daedalus*, summer issue, 1–17; Nakicenovic, N., 'Freeing energy from carbon', *Daedalus*, summer issue, 95–112.

[2] Demirdöven, N. and Deutch, J., 2004, 'Hybrid cars now, fuel cell cars later', *Science* **305**, 974–976.

[3] Angelini, A.M., 1977, 'Fonti primarie di energia', *Enciclopedia del Novecento*, Vol. II, p. 536.

[4] Pollack, H.N. et al., 1993, 'Heat flow from the Earth's interior: analysis of the global data set', *Reviews of Geophysics* **31** (3), 267–280.

[5] International Geothermal Association, 2001, *Report to the UN Commission on Sustainable Development*, session 9 (CSD), New York, April, p. 536.

[6] Munk, W. and Wunsch, C., 1998, 'Abyssal recipes II', *Deep Sea Research* **34**, 1976–2009.

[7] Isaacs, J.D. and Schmitt, W.R., 1980, 'Ocean energy: forms and propects', *Science* **207**, 265–273.
[8] Clarke, R.C. and King, J., 2004, *The Atlas of Water*, Earthscan, London, p. 45.
[9] Farinelli, U., 2000, 'Renewable sources of energy: potentials, technology, R&D', *Atti dei Convegni Lincei* **163**, Ac. Naz. dei Lincei, Roma, pp. 267–276.
[10] Jacobson, M.Z. and Masters, G.M., 2001, *Science* **293**, 1438. For some critical comments see DeCarolis, J.F. and Keith, D.W., 2001, *Science* **294**, 1000–1001, and response, *Science* **294**, 1001–1002.
[11] Cost estimates vary depending on allowances for intermittency, etc. In note [10] Jacobson and Masters suggest per kWh 3–4 (US) cents for wind, while the IEA is near 5–6 cents for wind in good sites and 35–60 cents for solar. Service, R.F., 2005, 'Is it time to shoot for the Sun?', *Science* **309**, 548–551 reports 5–7 cents for wind, 25-50 cents for solar photovoltaics, 2.5–5 cents for gas and 1–4 cents for coal, all per kWh. All such figures depend very much on what is included. In addition, how could one evaluate reliably the cost per kWh of climate change from coal generated electricity?
[12] Archer, C.L. and Jacobson, M.Z., 2005, *Journal of Geophysical Research* **110**, D12110, 1-20.
[13] Kempton, W. et al., 2007, 'Large CO_2 reductions via offshore wind power matched to inherent storage in energy end-uses', *Geophysical Research Letters* **34**, L02817, 1–5.
[14] Dresselhaus, M.S. and Thoma, I.L., 2001, 'Alternative energy technologies', *Nature* **414**, 332–337. The record 40.7% efficiency is from a news item in *Nature* **444**, 802, 2006; Lewis, N.S., 2007, 'Toward cost-effective solar energy use', *Science* **315**, 798–801.
[15] Dennis, C., 2006, 'Radiation nation', *Nature* **443**, 23–24.
[16] Marris, E., 2006, 'Drink the best and drive the rest', *Nature* **444**, 670–672.
[17] Sanderson, K., 2006, 'A field in ferment', *Nature* **444**, 673–676.
[18] Pimentel, D., 2003, 'Ethanol fuels: energy balance, economics and environmental impacts are negative', *Natural Resources Research* **12**, 127–134.
[19] Farre,l A.E. et al., 2006, 'Ethanol can contribute to energy and environmental goals', *Science* **311**, 506–508.
[20] Goldemberg, J., 2007, 'Ethanol for a sustainable energy future', *Science* **315**, 808–810.
[21] Himmel, M.E. et al., 2007, 'Biomass recalcitrance: engineering plants and enzymes for biofuel production', *Science* **315**, 804–807.
[22] Tilman, D. et al., 2006, 'Carbon-negative biofuels from low-input high-diversity grassland biomass', *Science* **314**, 1598–1600.
[23] Haag, A.L., 2007, 'Algae bloom again', *Nature* **447**, 520–521.
[24] Weinberg, A.M., 1986, 'Are breeder reactors still necessary?', *Science* **232**, 695–696.
[25] Klapish, R. and Rubbia, C., 2000, 'Accelerator driven systems', *Atti dei Convegni Lincei* **163**, Ac. Naz. Dei Lincei, Roma, pp. 115–135; Also Rubbia,

C., 1994, 'A high gain energy amplifier operated with fast neutrons', *American Institution of Physics Conference Proceedings*, p. 346.
[26] Bagla, P., 2005, 'India's homegrown thorium reactor', *Science* **309**, 1174–1175.
[27] www.iter.org
[28] Seewald, J.S., 2003, 'Organic–inorganic interactions in petroleum-producing sedimentary basis', *Nature* **426**, 327–333.
[29] Buffet, B.A., 2000, 'Clathrate hydrates', *Annual Review of Earth and Planetary Science* **28**, 477–507.
[30] IPCC, 2005, *Special Report on Carbon Dioxide Capture and Storage*, see Table 6.1.
[31] Beers, M.H. and Berkow, R., 1999, *The Merck Manual*, 17th edn, section 1, Merck Research Laboratories, Whitehouse Station, NJ.
[32] Grevesse, N. and Sauval, A.J., 2002, 'The composition of the solar photosphere', *Advances in Space Research* **30** (1), 3–11.
[33] Kerrich, R. 1999, 'Nature's gold factory', *Science* **284**, 2101–2102.
[34] Meadows, D.H. et al., 1972, *Limits to Growth*, Potomac Associates, London.
[35] Servan-Schreiber, J.-J., 1981, *Le défi mondial*, LGF, Paris.
[36] Goeller, H.E. and Zucker, A., 1984, 'Infinite resources: the ultimate strategy', *Science* **223**, 456–462.
[37] Rona, P.A., 2003, 'Resources of the sea floor', *Science* **299**, 673–674.
[38] Halfar, J. and Fujita, R.M., 2007, 'Danger of deep-sea mining', *Science* **316**, 987.
[39] http://minerals.er.usgs.gov/minerals/
[40] Chipman, A., 2007, 'A commodity no more', *Nature* **449**, 131.
[41] Albarède, F., 2003, *Geochemistry*, Cambridge Univ. Press, Appendix A.
[42] Elmsley, J., 2001, *Nature's Building Blocks*, Oxford Univ. Press.
[43] Teng, F.-Z. et al., 2004, 'Lithium isotopic composition and concentration of the upper continental crust', *Geochemica Cosmochemica Acta* **68**, 4167–4178.

8 The Future of Survivability: Water and Organic Resources

If you think in terms of a year, plant a seed; if in terms of ten years, plant trees; if in terms of 100 years, teach the people.

Confucius

8.1 Water

Water is perhaps the most essential commodity for life. In fact, much of the biosphere consists of water, and most processes in organisms involve the transportation of various compounds in watery form. Most plants consume water in liquid form and return it to the atmosphere as water vapor. Only a small part of our water needs is for human uptake, while our animals and, even more, our agriculture need far more. In the arid regions of the Middle East, where so many of our perceptions about the world were formed, water played an essential role in human relations; possession of water was a frequent source of conflict, while whole societies perished when climatic or ecological changes made water scarce. Today in the developed world when, for many people, water is just a commodity that comes out of a pipe, it has lost its value and the result is that much water is wasted unnecessarily [1]. However, since water availability is distributed very unevenly, there are many places where it is in short supply. Increasing numbers of people and increased *per capita* consumption then raise the question whether there is enough water in the world to further augment the consumption of what is after all a finite resource. Much concern on this issue has been expressed in recent times with confusion about water scarcity as such versus scarcity of clean water. Cleaning up water pollution, which is frequently due to poverty, carelessness or greed, should be feasible, while a real lack of water would be harder to cure.

Most water on the Earth's surface is in the oceans, but difficult to use for land organisms because of its high salt content. Less than 3% is fresh water, and lies mainly in the polar ice caps, with much of the remainder underground. The volume of fresh water in the world's lakes is some 170,000 km^3 (0.013%), nearly half of which is in the Caspian Sea.

8.1.1 The water cycle

The oceans continuously produce water vapor by evaporation. Much of this later condenses, leading to rainfall on the oceans. However, part of the water vapor drifts inland with the wind and leads to rain or snow on the continents (Figure 8.1) [2]. About two-thirds of this volume evaporate again or are transpired by plants and trees, while the remainder finds its way into rivers which ultimately dump the water into the sea. It is this water that can be used by humanity for domestic purposes, industry and agriculture (DIA). But the rain water that has fallen onto areas without irrigation and is later transpired by plants and trees, is also useful, as it may produce crops and meadows for animals. The river water is sometimes called 'blue water', while the water later transpired by plants is called 'green water'. World wide the annual flow of blue water amounts to some 40,000 km^3, the green water to somewhat more. Both flows are still rather uncertain. Evaluation of the green flow is tricky because one has to separate the water vapor produced by evaporation from that produced by transpiration. The green water is directly related to the natural biological productivity of the land, while the blue water lends itself to human manipulation.

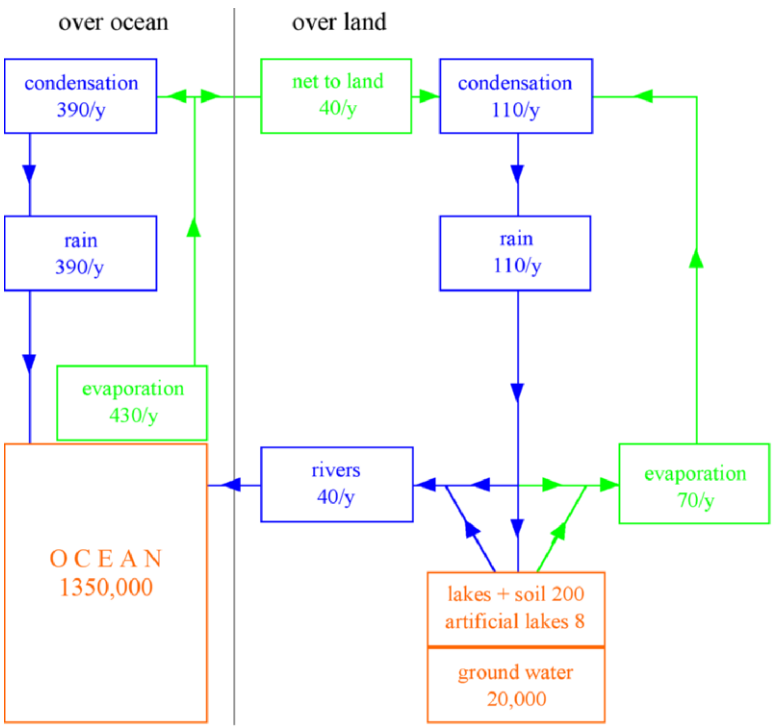


Figure 8.1 The hydrological cycle [2]. Lines in blue are liquid water fluxes, in green water vapor fluxes, both expressed in thousands of km^3 liquid water equivalent. Oceans, ground water, lakes and soil in orange indicate the volumes contained therein in thousands of km^3. Much of the ground water is located at considerable depth. Rain includes snow and evaporation includes transpiration by plants.

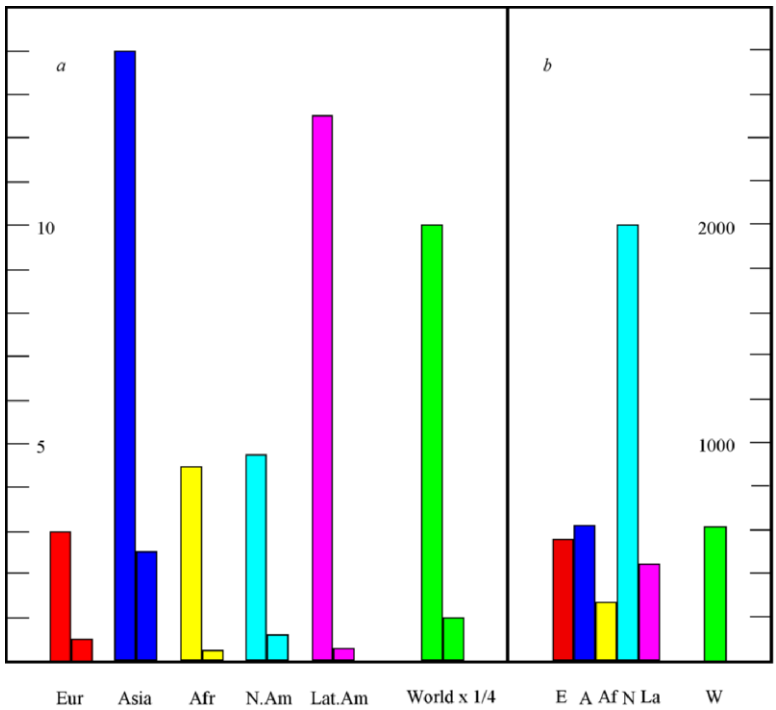


Figure 8.2 (a) River runoff (left bar) [1], [3] and water withdrawals (right bar) in units of 1,000 km^3/year in five continents, and for the world divided by 4. North America refers to USA/Canada and Latin America includes the Caribbean. (b) Per capita annual water withdrawals in the same five continents and in the world in units of m^3/year per capita [3].

Figure 8.2 indicates the quantities of river water available on the five well-populated continents. It is seen that the New World is more favored than the Old.

8.1.2 Water use and water stress

Figure 8.2 also shows the estimated (rather uncertain) water withdrawals by humans, which amount to around 10% of the available water. So one might wonder how there could be a water problem. Three factors are important: (a) the accessibility of the rivers; (b) the difficulty of collecting flood waters; and (c) the extreme unevenness of the water resources.

(a) Half of the river flow in South America is due to the Amazon, but few people live there, so, in some sense, that water cannot be used effectively for human purposes. The same is true for some of the rivers flowing into the Arctic and to a lesser degree for the Congo. It has been conservatively estimated that some 19% of the global runoff is inaccessible [1].

(b) In regions with a monsoon-dominated climate much of the rainfall occurs

in cloud bursts that may lead to flooding, but give only a limited benefit to year-round agriculture, unless the water is captured in reservoirs. Total reservoir capacity world wide is estimated at 7,700 km^3, to be compared with water withdrawal by humans of 4,000 km^3 annually [3]. Hence, even though some of the reservoirs are only used for power generation, it would seem that the non-flood flow plus the reservoir-based flow should suffice for *average* human needs. In fact, according to one estimate the geographically and temporally accessible runoff amounted to 12,500 km^3 per year of which 35% were withdrawn [1]. Actually, the overall situation may be still more favorable, since some of the water withdrawals could be re-used. For example, much of the 1000 km^3 withdrawn for industry is for cooling of power plants and subsequently is returned to the river flow. Also, a third of the irrigation water could be used a second time [1], although in practice it may have been polluted by insecticides.

In Table 8.1 we have assembled some data on annual runoff on the continents [3], the runoff corrected for inaccessibility and flood loss with more favorable assumptions than in [1], since more reservoirs have been made and more will be constructed, and on the annual withdrawals (from [3] updated to 2005 in proportion to the increase of the population). Note that in Figure 8.1 North America is only the USA + Canada, while in Table 8.1 North America is the continent down to Panama.

Table 8.1 Annual water runoff on the continents

Subsequent columns give Q, the total runoff per continent in km^3, and the same corrected for inaccessibility, the corrected value in m^3 per capita, the human withdrawals in m^3/per capita and the fraction withdrawn. All values are for estimated 2005 population numbers. Inaccessibility corrections are 6,500 km^3 for Amazon/Orinoco, 1,000 km^3 for Congo and 1,000 km^3 and 700 km^3 for the rivers that come out into the Arctic Ocean, respectively, in North America and Asia. The results are subsequently halved for the loss of much of the flood waters which is only partially compensated by reservoirs. At the present moment this may still be an optimistic assumption in Africa.

	Total	Corr	Corr	Withdrawals		Ratio
	km^3	km^3	m^3/cap	km^3	m^3/cap	
Europe	2,800	1,400	1,900	430	590	0.31
Asia	13,700	6,500	1,700	2,300	600	0.36
Africa	4,500	1,750	2,000	250	270	0.14
North America	5,900	2,450	5,000	750	1,560	0.31
South America	11,700	2,600	6,800	140	400	0.05

(c) While, therefore, globally there is no water shortage, the very uneven distribution leads to serious problems. When agriculture began, people settled in the most fertile regions: the valleys of the Nile, the Babylonian rivers, the Yellow river and others. Water was not a problem and the population rapidly expanded,

after which it became a problem. A further recent factor has been fertilizer-based agriculture, which increased productivity but required a large volume of water. As long as the population was not too large people could move when environmental conditions deteriorated, but this has become much more doubtful when hundreds of millions of people are involved.

How many people have water problems? One common definition of water stress is when water withdrawals for domestic, industrial and agricultural use (DIA), exceed $0.4Q$, with Q denoting the available water [3]. The UN at first evaluated DIA/Q on a country by country basis and found in 1995 that the number of people living in water-stressed areas (i.e. DIA/Q, > 0.4) was 460 million. However, country averages are deceptive because of the difference inside large countries. China as a whole is not water stressed, but several hundred million people in northern China are. So a more meaningful result was obtained by evaluating DIA/Q without looking at national boundaries, on a grid of 59,000 squares, 0.5×0.5 degrees in longitude and latitude – i.e. about 50×50 km. It was found that instead of 460 million people on the country by country basis, actually 1,760 million people lived in a square with DIA/$Q > 0.4$ and were therefore considered to be water stressed [3]. An alternative measure of water stress counts the number of persons who have to manage with less than 1,000 m^3 of water per year [4]. In a way this seems to be a more objective measure than DIA/Q which makes people who waste much water seem to be water stressed.

8.1.3 Remedial measures

Since it would be difficult on a large scale to move people to where the water is, we should explore the possibility of moving the water to where the people are. Could we increase the water supply in the drier areas? Five possibilities may be identified:

- deflecting the water from water rich to drier areas;
- building more dams to capture flood water;
- utilizing ground water;
- moving more of the agriculture to where the water is;
- desalinating ocean water or local water with a high mineral content.

We now review these different options.

Plans for deflecting rivers to dry areas

During the sixties two gigantic projects were studied which would have deflected remote Arctic rivers to the dry zones further south. In North America this would have involved the transfer of the water of the Yukon and other rivers to California and other dry regions in the western USA. Various possibilities were considered for the transport of the water – huge pipes crossing some of the Rocky Mountains, or even large bags filled with water being tugged over the Pacific. Remarkably, many Americans thought that the water would be for free since it is heaven's bounty and were surprised that the Canadians did not necessarily agree!

Equally gigantic were the plans in the USSR for tapping the waters of Ob–Yennesej–Irtys and sending these down to the dry regions east of the Caspian Sea, forming there a lake of more than 200,000 km^2 – half as large as the Caspian Sea. Neither of these two projects was realized because of its high cost. Also the ecological consequences of the creation of such a lake in Siberia were far from obvious. While these projects have never been executed, a mega project in China is currently under construction. When completed in 2050, it would transport some 45 km^3 annually from the Yangtse to the Yellow river. The first stage, the construction of the Three Gorges Dam, 1,500 meters wide and 180 meters high, should be completed by 2009 [5]. Behind the dam a reservoir with a capacity of 39 km^3 would generate some 18 GW of electrical power. The 45 km^3 of water could give 300 million people on average 150 m^3 of supplementary water.

Building dams

Numerous dams have been built in the world to collect flood waters for agriculture and to generate hydroelectricity. The worldwide volume of the reservoirs amounts to some 7,700 km^3 and is still increasing. Irrigation and power production have benefited large numbers of people and have contributed much to the world's ability to feed 6 billion people. However, there are also serious problems.

For the reservoirs behind the Three Gorges Dam nearly 2 million people had to be evacuated. In many other cases thousands or even hundreds of thousands have had the same fate. Especially in densely populated areas, resettlement of the people is difficult, since all usable land is already occupied. Furthermore, the whole river environment may be negatively affected. An example is the Assuan Dam in the Nile river. Because of the dam all the silt that, in the past, fertilized the surrounding areas is held up. Moreover, in the Nile delta erosion has become a serious problem because no new material arrives, while also the fisheries there have suffered [6]. The same is also the case at the mouth of the Yangtse [7]. In tropical areas stagnant water creates the ideal environment for a variety of diseases, and dam construction has sometimes had negative health effects [8]. Finally, the quality of dam construction is essential, especially in areas with earthquakes, as the breaking of a dam creates sudden catastrophic flooding far downstream.

Therefore, a very careful weighing of all the consequences of the construction of a dam is necessary. At the same time, it is all too easy for the people in the developed world to criticize the dam construction in the less-developed countries. If in northern China several hundred million people have inadequate water, is there a realistic alternative to the diversion of the waters of the Yangtse? In the developed world one can afford to even destroy some existing dams to restore a more optimal river ecology. But elsewhere the choices are harsher.

Utilizing ground water

Ground water appears to be abundant (Figure 8.1); however, much of that water is at great depth. In practice, it has appeared that in many of the acquifers that

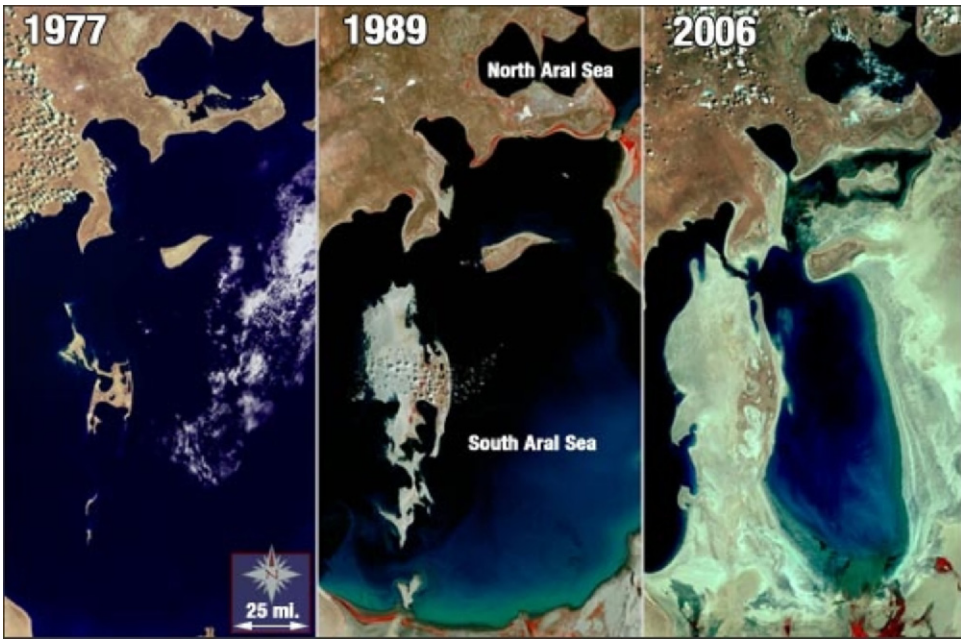


Figure 8.3 Evolution of the Aral Sea. (*Source*: NASA.)

Box 8.1 The Aral Sea

Situated 600 km east of the Caspian, the Aral Sea was the world's fourth largest lake. Fisheries prospered and surrounding areas were irrigated with a modest amount of water from the rivers flowing into the lake. From around 1960 the water withdrawals increased to satisfy the requirements of great cotton growing projects in central Asia. As a result, the lake lost much of its volume and salinity reached oceanic values. The worst aspect was that a large part of the salt sediments at the lake bottom was exposed. Subsequently, increasing storms spread the poisonous dust over large areas. The fisheries ceased and even drinking water became affected, with serious consequences for the health of the population in a large area [9]. Fortunately, recent attempts to undo the damage by adding water to the lake appear to have had some success [10].

have been exploited, the water level has gone down rather quickly. As examples, in Chicago the water table had gone down by 274 meters by 1979, while in the North China plains it is falling by 3 meters each year [11]. In coastal areas a sinking water table leads to an inflow of salt water. So, while there may be particular areas on the globe where it is reasonable for some time to 'mine' the

ground water, it cannot be a sustainable solution for long, and we have to live with the water that rivers provide. Even less sustainable is the mining of lake water that is not renewed by rivers. In fact, it may lead to major disasters of which the Aral Sea (see Figure 8.3 and Box 8.1) is the prime example. Another problem is the high mineral content of some of the ground water. In Bangladesh, where the surface water was rather polluted, numerous wells were made to tap the deeper ground water. After some time it was discovered that many people suffered from arsenic poisoning. The arsenic in the water is entirely of natural origin, related to volcanic activity long ago. Further analysis has shown that some 40 million people in Bangladesh and West Bengal are at risk [12]. Finally, extracting water at great depth takes a lot of energy; in fact, bringing the water up from a depth of 500 meters takes as much energy as obtaining it from the desalination of sea water.

Agriculture where the water is

In a fully rational world one could envisage moving the agriculture to where the water is and then transporting the resulting products to where the people are – the so-called 'virtual water' trade. Here and there this is actually taking place today with the Brazilian cultures of sugar cane and soybeans rapidly increasing. However, this could work only if the people who are currently engaged in agriculture elsewhere could be employed in industry in order to have something to trade, which it hardly seems possible for the moment.

Desalination

The desalination of sea water may make an important contribution. In the past it was very expensive owing to the energy required, but with more modern technology it has become much more affordable (see Box 8.2). To gain an idea of the order of magnitude of the energy required, the production of 40 km^3 of fresh water, which is about equal to the amount of water provided by the Three Gorges Dam, requires less than 70 TWh of electrical energy – half as much as the power produced by that dam.

8.1.4 Water for 100,000 years

We assume again that the world population stabilizes at 11 billion people. We adopt a minimum water requirement of 1,000 m^3 per year per capita [4, updated]. At present, of the major countries only the USA has significantly larger withdrawals than that [13]. So the total global annual requirement becomes 11,000 km^3, nearly three times present-day global withdrawals. We have previously quoted the estimate of accessible runoff, including that captured by reservoirs of 12,500 km^3 per year [1]. This was based on a storage capacity of 5,500 km^3 in reservoirs of which only 3,500 km^3 were used for regulating river runoff. In the meantime, the reservoir capacity has increased to 7,700 km^3 [4]. Upon adding the increase to the previous figure we come to an accessible runoff figure of 14,700 km^3 annually. Assuming that still a further modest 1,300 km^3 of reservoir volume will be added in the future and that we prudently use only half

Box 8.2 Desalination

The removal of salt and chemical or bacteriological pollutants is not very difficult, but in the past it took an inordinate amount of energy. The simplest procedure is distillation by boiling the water and later condensing the water vapor. To evaporate 1 m^3 of water takes about 700 kWh of heat energy. While that energy can be partly recuperated during the condensation, the process is only really feasible in places with very cheap energy or very expensive water. However, the newer procedure of reverse osmosis is much more economical. The salt water is pushed at high pressure through a filter that lets the water molecules through, but not the salt and other impurities. The development of efficient filters that do not clog up, and of schemes that also recycle the pressure, have continued to reduce the power required. A commercial plant on Belle-Île, an island in front of the French coast, should arrive at 3.2 kWh of electrical energy per m^3 of purified water [14], while in California with lower pressure membranes, 1.6 kWh/m^3 was reached [15].

When salt is dissolved at oceanic concentrations in pure water, 0.8 kWh/m^3 of heat energy is released and so this is also the minimum energy required to remove it again. Thus, technological improvements seem possible to further reduce the energy costs [15]. But even at 1.6 kWh electrical energy per m^3, the cost would be no more than 1,600 TWh of electrical energy to desalinate 1,000 km^3 of ocean water, equal to one-quarter of current worldwide water consumption. If the needed electricity were generated by thermal plants, about 15 EJ of heat energy would be required, corresponding to less than 4% of present-day energy consumption. But, of course, it would be far more rational to use electricity generated by wind or by solar panels. Less than 5,000 km^2 of panels at 15% efficiency would suffice to produce 1,000 km^3 of water in tropical deserts. Intermittency of the solar energy would not be that much of a problem if limited storage facilities for the water would be built. The other advantage is that there is no need for gigantic facilities, since the plants could be scaled to local needs.

The desalination of 1,000 km^3 of water implies the production of some 30 km^3 of salt or perhaps more probably something like 100 km^3 of salty brine. While some of this may be used for other purposes, most will have to be disposed of in ways that do not have too much effect on the local salinity of the ocean.

of the resulting 16,000 km^3 in order to leave ample water in the rivers, we would need to find an additional 3,000 km^3 of water per year. If we could obtain that by desalination by the most efficient methods currently available, that would cost about 0.5 TWy of electrical energy, which is 1% of the projected 63 TWy of expected future energy use (see Section 7.1). Evidently other scenarios for the repartition of water use over rivers, reservoirs and desalination are entirely

possible. But if the projections of future energy use are realistic, the extra energy cost is acceptable. Since some 60% of the world population lives less than 200 km from the seashore, the use of desalinated water could solve many problems. So, if the world developed its renewable energy sources, water problems would no longer occur in most areas. In the long term, enough water appears to be assured in the world. However, since the development of an adequate energy infrastructure will probably take a century or more – and as large population increases also appear imminent – the water situation for the next 50–100 years looks much less favorable.

8.1.5 From now to then: water and climate change

Long term there is no water problem as long as enough energy is available. However, the present situation is more difficult since even the minimal energy needs are not satisfied in much of the developing world. So, large-scale desalination is not an option there. At the same time, population pressures in Africa and parts of Asia will increase, some aquifers will go dry and climate change will gain in importance. Not only will global and regional changes in precipitation occur, but the stabilizing effects of glaciers, forests and wetlands are also likely to diminish. With increasing temperatures the hydrological cycle will speed up because of increasing evaporation and water vapor in the atmosphere. Nevertheless, climate models generally predict increasing drought in the northern and southern parts of Africa, in Central America, Australia and in southern Europe [16]. Equally worrying is uncertainty about the Asian summer monsoon. While some intensification seems probable, various models predict that the interannual variability is likely to increase [17]. Climate models tend to suggest some improvement in rainfall in northern China and Pakistan.

Various scenarios have been made of population and economic development (Chapter 6). Here we consider scenarios B1 and A2. Scenario B1 corresponds to rapid development, implementation of clean energies and a world population that peaks in mid-century, and it is the most favorable of the scenarios usually considered. A2 corresponds to slow development and continuing population increases, and is one of the most pessimistic scenarios. For these scenarios, models of the projected water stress in 2075 have been based on criteria of people living in areas with <1,000 m^3 per capita or, alternatively, with withdrawals over runoff DIA/Q >0.4 [4]. The results in Table 8.2 show that even in the most optimistic scenario, the number of people living in areas of water stress will still double by 2075, while in scenario A2 it could quadruple, mainly as a consequence of the population increase.

In tropical and wet parts of Africa rainfall may increase, and parts of the Sahel would benefit from a northward movement of the monsoon. However, both the north and the south should become substantially drier. Since the relation between rainfall and perennial runoff is quite non-linear, even 10–20% decreases in the former may lead to much larger reductions in the latter. For much of South Africa, Zimbabwe, Zambia and Angola, recent studies suggest that the remaining perennial runoff in 2100 will be no more than 65–80% of present values, and for

Table 8.2 Projections of numbers of people living, in 2075, under conditions of water stress.

Projections of numbers of people living, in 2075, under conditions of water stress. Subsequent columns give the scenario (see Table 6.2 and preceding discussion), the world population in billions, the CO_2 concentration, the global temperature increase from the year 2000, the numbers of people N (in billions) with less than 1,000 m^3 of water per capita per year and their percentage of the total population. All figures for scenarios B1 and A2 pertain to the year 2075. For comparison in the last line the corresponding figures are given for the year 2000.

Scenario	Population	CO_2 (ppmv)	ΔT (°C)	N (<1,000 m^3/(cap, yr)	%
B1	7.8	527	1.55	2.8	36
A2	13.2	659	2.4	7.0	53
2000	6.1	368	0	2.3	38

Morocco, Algeria, Tunisia it will be 0–50% [18]. While regional climate models still have much uncertainty, such predictions are a cause for concern.

Another danger of global warming is the melting of the permanent ice caps and glaciers on tropical mountains. Even in areas in the plains where there is no rain during the hot season, the annual melting of the ice provides a continuous runoff. During the wet season, snow then restores the ice. Increasing temperatures first cause the runoff to increase as the ice melts, but once it is gone, summer droughts become severe. The ice in the tropical Andes melts rapidly because of greater warming at 6,000 meters than lower down [19]. For example, in Peru the glaciated area has been reduced by 25% in only 30 years [20]. Still more worrisome, China, India and other countries obtain much melt water from the Himalayan and other glaciers, but these glaciers are now retreating – for example, the one that feeds the Yangtse has retreated by 750 meters in the last 13 years [20]. Again in the Himalayan area, glacier melt runoff appears to have increased by one-third. When many of these glaciers have vanished, the amount of summer runoff will crash rapidly. Since, in the Ganges, 70% of the summer flow comes from melting glaciers, the future looks uncertain [20]. The solutions to such problems require time, and some decades may still elapse before the problems become acute [20]. What can be done, other than building more reservoirs, is not evident.

8.2 Agriculture

8.2.1 Increasing productivity

Effective domestication of plants and animals began some 9,000–10,000 years ago in the Middle East and in China around the time of the end of the Younger Dryas (see Chapter 5), which terminated the last ice age [21]. This allowed the production of much more food and permitted a rapid increase in the population. Climate changes, population pressures, hostile neighbors and invaders caused populations to move. This, and trade, led to the spread of agricultural knowledge

to other regions, some of which had even better conditions – more productive soils and sufficiently ample and reliable rainfall. Gradually, skills in water management were also developed: canals were dug that could dispose of excess water, and irrigation channels for periods of shortage. Unfortunately, the increasing food production also led to large increases in the human population. Much in the way described in 1798 by Thomas Malthus in *An Essay on the Principle of Population*, population grew rapidly to just beyond the maximum that could be properly fed [22]. While Malthus has been much maligned, the essential correctness of his argument can hardly be denied; only effective population control has allowed the food supply per capita to be maintained in the long term.

Two factors contributed to the next jump in food production: the mechanization of agriculture and the development of the chemical industry during the last century. The former reduced the manpower needed to farm a certain area, while the latter generated efficient fertilizers and insecticides. Somewhat earlier a wholesale exchange of plants and animals between the old and the new world had produced a greater variety of agricultural products in both. Domesticates grown far from their places of origin include Mexican maize, tomatoes and potatoes in Europe and European wheat, cows and horses in the Americas.

Increased understanding of plant and animal physiology and genetics allowed substantial improvements in the productivity per plant or animal. Thus, present-day rice, maize and wheat plants produce significantly more cereals per plant than their predecessors, while cows are bred with much increased milk production. The last step in these developments has been the appearance of genetically modified (GM) plants and animals in which entirely new genes from other organisms are inserted to obtain desirable characteristics, such as resistance to certain pests, herbicides and drought or increased content of nutrients that result in better yields and water utilization. These later developments have not yet been generally accepted, and a strong antagonism against GM foods has developed in many countries. While this may be due in part to real fears about their safety – for example, concerning the diffusion by the pollen of the modified genes – the aggressive way in which some US firms (the main world seed groups) have tried to force these foods on the consumer have greatly contributed to the hostility. After all, food is too important to be treated as just any industrial commodity.

Agriculture has also brought a number of problems in its wake. To make room for crops, deforestation was frequently necessary. When done excessively, unnecessarily or in inappropriate locations, it leads to erosion and damage to soil, forest and river systems. The application of excessive amounts of fertilizer leads to algal blooms which destroy aquatic life. As an example, at the mouth of the Mississippi, an area of 20,000 km^2 in the gulf of Mexico has lost almost all of its life, except for algae which suffocate everything else [23]. Many insecticides are not sufficiently selective and so damage entire ecosystems. The classical case is DDT (used against mosquitoes) which caused the demise of many species of birds (and others), so eloquently described in 1962 by Rachel Carson in her influential book *Silent Spring* [24]. Nevertheless, it has to be acknowledged that keeping damaging insects in check is necessary. Malaria spread by mosquitoes is

killing more than a million people a year in the tropics and for the moment a modest use of DDT appears necessary to reduce this.

Adequate production of food is the prime requirement for humanity. Owing to the development of modern agriculture it has been globally achieved, for the moment. The present worldwide agricultural production could easily suffice to feed the 6 billion people that inhabit the Earth [25]. However, some regions have serious shortages, while, in others, farmers are paid to produce less. Agricultural subsidies in the richer countries further distort the situation. Hence, the insufficiency in part of the world is more a question of economic organization. If there were a sufficiently strong will to deal with the problem, it certainly could be solved. With food production becoming a smaller and smaller part of economic activity, this would not even need to have a significant impact on the world economy. However, it would be meaningless if not accompanied by measures to limit population growth in the countries that currently have inadequate food production. Recently the combination of high oil prices and the development of biofuels are posing additional risks to the world food supply.

8.2.2 Present and past land use

The area of the Earth's land amounts to 150 million km^2, of which an uncertain 4 million km^2 are occupied by lakes and rivers and 16 million km^2 by ice, leaving an accessible land area of 130 million km^2. Of this land some 31.5% is currently covered by forest, 32% by grass and shrubs (of which 26% is used for pasture), 11.5% is crop land, 5% tundra and 20% desert (Figure 8.4) [26–29]. Thus 37.5% of the land area is directly used by humans, to which should be added the part of the forests that is periodically harvested.

In the course of time the forest area has diminished as the agricultural area has increased. By some estimates forested area diminished by 19% between the years 1700 and 1980 [28]. There is some evidence that in recent decades the decline in forests has been halted or even reversed, at least in the temperate regions where much wood had been harvested for timber and firewood in the previous century.

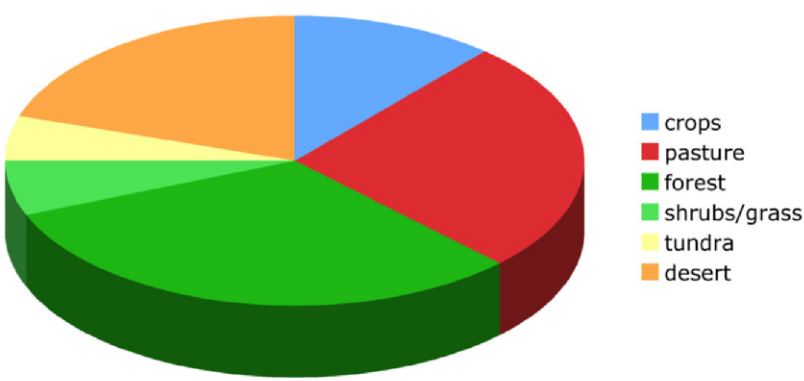


Figure 8.4 The distribution of the accessible land use.

However, in the tropical forests much wood is being lost, though quantitative figures are quite divergent between different studies, while governments may have political motivations not to produce accurate statements. The agricultural area appears to have been stabilizing during the last decade with increases in the less-developed countries being compensated by decreases in the developed world. This is a consequence of the great increases in agricultural yields which have, in most areas of the world except Africa, exceeded on a percentage basis the increases in population. In the following sections we shall consider somewhat more quantitatively the various factors that have affected agricultural needs and performance.

8.2.3 Population

The world's population in 2005 amounted to some 6,500 million people, some seven times larger than estimated for the year 1800 [27]. Increases were even more rapid in the developing world: a factor of nearly 11 in the countries that constituted, in the past, British India and a staggering factor of 26 in Indonesia [30]. While such increases posed problems for the food supply, they also had a very negative effect on the social structure in agricultural production [31]. For example, the inheritance laws led to the land being split into ever smaller parcels, which ultimately produced a large class of landless agricultural laborers with low per capita productivity.

In the developing countries, the rate of increase over the last five years has come down to some 1.5% per year from 2.2% per year 30 years earlier [29]. At such a rate their present 5,000 million inhabitants would still nearly double by 2050. The reason the situation is not worse is mainly due to the rigorous population policies in China. For the other less-developed countries (LDCs) the current rate of increase averages 1.75%.

8.2.4 Agricultural land and production

The global areas devoted to crops and to pastures are currently both almost unchanging with an annual increase of a few tenths of a percent in the LDC being compensated by a corresponding decrease in the developed world [29]. The increased yields have globally kept pace with the increased population. Cereals (maize, rice, wheat and less important others) are at the basis of much of human nutrition, either directly or less efficiently through animal feed. It takes some five calories of cereals to make one calorie of chicken [32] and more than that for other meats. In Table 8.3 we have assembled some data [29] about cereal production per hectare and per capita in different parts of the world, and about the corresponding changes over the last 30 years. It is seen that, except in Africa, the cereal yields have nearly doubled and per capita the typical harvests have gone up by some 15%. The somewhat lower values in India are associated with a more vegetarian diet. The low values of African per capita cereal production are worrisome, even though Africans in the tropics consume a wider range of other foods. What appears to be particularly serious in the cereal production is its low growth rate and its decline per capita. The reasons will become clear below.

Table 8.3 Cereal production in 2005 and 1975

Subsequent columns give the geographical areas (LDC = Less Developed Countries only), the population in millions (extrapolated from 2004), the cereal production in millions of metric tons, the corresponding crop area in millions of hectares, the yield in tons per hectare and the cereal production in kilograms per capita, all for the year 2005. The last four columns give the ratio of the 2005 figures to those for 1975 for the population, the crop area, the yield and the production in kilograms per capita. It should be emphasized that roots and tubers provide additional nutrition, especially in many parts of Africa. All figures are based on the early 2006 data from the FAO, the UN Food and Agricultural Organization. Note that all of Africa (except South Africa), the major Asian countries (except Japan) and all of Latin America are considered as LDCs. In reality, different countries in the 'less-developed world' are at very different stages of economic development.

						ratio 2005 over 1975			
Area	**Mpop**	**Mton**	**Mha**	**ton/ha**	**kg/cap**	**pop**	**ha**	**ton/ha**	**kg/cap**
Africa (LDC)	842	116	97	1.2	138	2.2	1.7	1.19	0.92
China	1,330	428	83	5.16	322	1.43	0.85	2.07	1.23
India	1,097	236	100	2.37	215	1.77	0.98	1.88	1.04
Other Asia (LDC)	1,280	362	117	3.09	283	1.89	1.18	1.81	1.13
Latin America	558	158	51	3.13	284	1.73	1.02	1.92	1.13
Developed World	1,345	940	239	3.94	698	1.19	0.78	1.77	1.16
Total World	**6,452**	**2240**	**686**	**3.27**	**347**	**1.59**	**0.96**	**1.71**	**1.04**

8.2.5 Irrigation

In Asia about one-third of the agricultural land is being irrigated, while in Africa the figure is close to 5% and in the rest of the world 11% [29]. Irrigation would be highly desirable in Africa where large tracts have low rainfall. Its benefit is seen from the fact that 40% of the world's food comes from the 16% of agricultural land that is irrigated [25]. Irrigated land has been increasing globally by some 1.8% per year over the last 30 years [29]. It is mainly limited by the need for reservoirs and dams, the construction of which frequently tends to displace many people in rather densely populated areas. Other problems are the silting up of the reservoirs with the particulate matter carried by the rivers that fill them and the tendency to salination in waterlogged areas where the evaporation of the water lets salts emerge from below.

8.2.6 Fertilizers and pesticides

Some soils are poor in available mineral nutrients. Others become so after harvests and the runoff from irrigation have removed them. The two most important fertilizing elements are nitrogen and phosphor. Most of the atmosphere (78%) is composed of nitrogen (N_2), but most plants are unable to metabolize it directly. Instead they obtain their nitrogen from nitrogen compounds in the soil. Some success may be obtained by growing trees or

plants that can process the atmospheric nitrogen in between the plants that cannot, but for high-yield agriculture it is necessary to add nitrates or ammonia-based fertilizers. Huge supplies of nitrates were found in Chile and Peru, but were superseded by the process invented by Fritz Haber in 1908 which allowed the synthesis of ammonia from atmospheric nitrogen [33]. Once this step had been taken, the appropriate fertilizers could be produced industrially. The nitrate mines in northern Chile were abandoned, which caused much social upheaval. As the fabrication process of ammonia takes a fair amount of energy, the presently increasing costs of oil have given new life to some of the mines. Phosphor is mined in the form of phosphates which may be used directly. Particularly rich deposits occur on the coast of the Sahara desert, and this has caused continuing problems between Morocco and the inhabitants of the region. Nitrogen is an essential component of chlorophyll, the molecule responsible for photosynthesis in plants, and of DNA; phosphor is a component of the molecules of DNA and ATP, the energy provider for synthesis reactions. Many other elements are needed but are usually adequately available in soils. However, potassium supplements are also frequently required and occasionally small amounts of elements such as copper [34].

Pesticides are an important part of modern agriculture, notwithstanding their environmental problems. Unfortunately new pesticides have to be developed continuously, because resistance develops in a small number of years [25]. Until now temporary success in the fight against pests has generally been obtained, but examples such as the Irish potato famine in the middle of the 19th century, or the grape phylloxera in Europe somewhat later, illustrate the dangers. Probably the only way to guarantee food security is to avoid monocrops; biodiversity is effective, because many insects and diseases are species specific. Insecticides are also essential to avoid losses during transport and storage. It has been estimated that in 1960 almost a third of the Asian rice harvest was lost to insects.

8.2.7 Top soil

Fertilizer may compensate for the loss of nutrients. However, soil is more than just chemicals. Organic matter is a particularly important component of top soil [35], as it and the biota, like worms and others which live in it, improve the physical characteristics of the soil needed for the growth of plants [36]. The loss of soil has been well documented in the USA where, in Iowa, half of the top soil has been lost during the farming activities of the last century and a half [36]. In China the fertile loess areas are eroding rapidly, creating nowadays damaging loess storms in Beijing and elsewhere [37]. Remedying such losses is a difficult but necessary task.

So, the overall conclusion would seem to be that current agriculture is globally able to feed the current world population, even though in some parts serious shortages occur owing to particular circumstances. Increased availability of fertilizers in Africa could probably lead to rapid improvements, but with the world population still set to nearly double, the constraints and the risks will become ever greater.

8.2.8 Agriculture for 100,000 years

Can the world in the long term feed 11 billion people? At some level of well-being the answer is probably yes, at least for some time. As was written in a recent review [25]: 'There is a general consensus that agriculture has the capability to meet the food needs of 8–10 billion people while substantially decreasing the proportion of the population who go hungry, but there is little consensus how this can be achieved by sustainable means.'

As an example take China, which produces 320 kg of cereals per capita annually based on a yield of 5.2 tons/ha. In the 11-billion-people world some 9,600 million would live in what are now the developing countries. If the area of cereal production remained unchanged, it would require 6.9 tons/ha to give them the same diet as China enjoys today. But the developed countries consume a lot more, some 700 kg per capita per year. If they then were to be fed at that level, a yield of 15 tons/ha would be required. If the production were to be spread uniformly over the entire world's cereal land, just somewhat over 11.2 tons/ha would be needed. What yields can be reasonably expected?

Much work is going on with Genetically Modified Organisms (GMO). This was mainly driven by the wish to make agricultural production in the developed world still cheaper than it already is, and to create monopolies on the most prolific seed strains. At the same time that this was done (in 1999), in excess of 60 billion euros in farm supports for cereals alone (out of some 220 billion for all agriculture) have totally distorted the trade, instead of increasing the production where it is most needed [25]. In a recent review the revealing comment was made that in China scientists could work on a wider range of GMO, because they worked in governmental service: 'As such they can undertake research on crop technologies that may be difficult to protect from the perspective of intellectual property rights' [38]. Could it be that the time has come to reform a patent system that leads to such results?

Present-day cereal yields in China and Japan average between 5 and 6 tons/ha. With further hybridization technologies combined with genetic modifications, Chinese agricultural scientists believe that ultimately rice yields up to 15 tons/ha could be reached [38]. Rice is a C3 plant which, in photosynthesis, first produces compounds with groups of three carbon atoms. C4 plants, which evolved only some tens of million years ago, are more efficient in the carbon fixation. One idea is therefore to convert rice into a C4 plant which, if successful, could alone increase yields by 50% [39]. It is thought by the scientists involved that after demonstration of the possibility it will take 12 years and US$50 million to develop the C4 rice. Is it not pathetic that such a sum could be an issue at a time that the OECD countries spend some US$220,000 million annually on agricultural price supports?

Of course, all of these improvements are still hypothetical. So let us look at the top yields that have actually been attained. Of the cereals, maize (a C4 plant) has a particularly efficient photosynthesis. In 2005 in Germany, Italy and Spain the yields all exceeded 9 tons/ha, reaching 12 tons/ha in Holland [29]. In the USA in 1992 the winning farmer in a maize-growing contest produced 21 tons/ha on a

4-ha plot, and several others came close to that [32]. But also, wheat yields in France, Germany, Ireland and the UK attained 7 tons/ha with Holland again at the top with 8.7 tons/ha. So, taking all the evidence together, it seems that there are no biological limits that would stand in the way of reaching the 7–15 tons/ha of cereal needed to feed the world with a more or less generous diet.

Of course, as all of these record harvests were obtained in the temperate zones, in conditions of good soil, water and fertilizer, there is no certainty that this can be reached everywhere. However, it is encouraging that experiments on soil with the characteristics of Amazonia grew yields of 7 tons/ha for a variety of crops over 17 years [40].

Would there be enough fertilizer, water and soil? Nitrogen-based fertilizers should pose no problems as long as the necessary energy is available to make them from the atmospheric gas. Phosphate is a different matter. Currently 30 million tons (Mt) annually are used, but total consumption would become at least 50 Mt if all agriculture would be provided with the presently typical amounts per hectare in the developed world. In a period of 100,000 years, the total would be 5,000 Gt to be compared with a presently estimated reserve base of 33 Gt. This could be a problem, as we noted in Section 7.2. We explored there the possibility of processing annually around 10 Gt of crustal rock. Since phosphorus is a rather abundant element (0.1%), this would produce 10 Mt of phosphorus. With phosphate slightly less than half composed of phosphorus, some 20 Mt of phosphate would be obtained annually. With a more efficient recycling of agricultural and animal wastes, this would probably be sufficient. This is true *a fortiori* of potassium, which still has a 21 times higher abundance in the Earth's crust.

Would there be enough water? If we wish to achieve the high yields necessary to feed 11 billion people generously, we probably would have to irrigate all the land currently used for raising cereals. At 12,000 m^3/ha [41] this corresponds to some 1,200 km^3 in Africa and 3,600 km^3 in Asia or then per capita 540 m^3 annually. This would be doubled on account of the irrigation of other crops and reduced by a quarter on account of normal rainfall to some 800 m^3. However, by implementing precision irrigation in which the water is provided when and where it is required, sufficient economies in water use might well be achieved to satisfy all other needs within the adopted 1,000 m^3 (Section 8.1.4).

In general it may be expected that the evolution of sustainable agriculture will proceed with many small steps adapted to the local soil and climatic conditions, perhaps in addition to a few more major steps relating to genetic modifications. From the present point of view there is no reason to believe that food would be inadequate to feed the 11 billion people well for 100,000 years, but the numbers indicate that we are rather close to the upper end of what is possible, and so extreme care will be needed along the way.

If we succeed in attaining the yields that seem possible, we may be able to keep the agricultural area to its present dimensions. This would then also imply that we may preserve the forests and perhaps even the wild areas of the world. One of the great uncertainties is associated with the changing climate. If the ice caps of

Greenland and west Antarctica were to melt, sea level would be raised by some 12 meters (see Chapter 6) and much of the particularly fertile land in the river deltas would be lost.

Of course, agriculture is more than just cereal production. But since cereals are the basic foodstuff for much of humanity, we have concentrated the discussion here. Moreover, such a wide variety of fruits, nuts, roots and vegetables are consumed in different parts of the world that it would be difficult to avoid going into too much detail.

8.2.9 From now to then

It seems likely that humanity can grow enough basic foods on present-day acreage and thus can spare enough land for nature. However, in the coming 50 years populations will still double in various parts of the world. This will strain the immediate food supply and could easily lead to a great push to open new agricultural land, especially in Africa. Even if later we can survive with present-day agricultural areas, people are not going to survive the 50 years of still rapid population growth combined with inadequate yields, and there will be a strong push to cut down the forests even though that soil is poor. If we value the long-term survival of Nature, forests, animals, etc., we have to find a way to make the agricultural productivity rise more rapidly in Africa and other areas. Probably the provision of fertilizer more or less free of charge could give a major push to productivity. If this were made contingent upon the effective protection of nature areas, success might be obtained. But since the population of Africa has doubled in only the last 26 years, this will be a now or never opportunity.

In this connection an interesting story appeared recently [42]. The African country Malawi had been for years a disastrous case depending upon food aid. The World Bank had tried to pressure the country to adhere to free market policies and not to subsidize agriculture, a curious policy if one sees the magnitude of such subsidies in the developed countries. In 2005 a new President decided to disregard all advice and amplified subsidies for fertilizer. As a result, corn harvests doubled in 2006, tripled the following year, and thereafter a surplus could be exported.

8.3 Forests and wilderness

A million people visit Lake Louise, a rather remote lake in the Canadian Rockies, to admire the magnificent scenery and re-establish their ties with the land. Many young people track through the woods and mountains to have a more intimate participation in Nature. We humans came all as a part of Nature, and it is only some millennia that we have begun to separate ourselves off and live in more artificial man-made environments. It is not surprising then that our atavistic instincts induce many of us to seek Nature when we have the opportunity. Some will find Nature on the oceans and most of us probably in the woods, savannahs and tundra lands. For many the Nature or wilderness

experience relieves the daily stress and thus makes an important contribution to our mental health.

Nowadays there is much talk about the need to protect biodiversity, and this need is justified with the belief that, in the jungles of the world, useful products for medicine may be found. But probably for most of us the issue is much deeper than the utilitarian aspects. The preservation of Nature has an intrinsic value. Most of us will never see a polar bear. Nevertheless, it is satisfying that they still exist in the wild regions of the world. Their loss would diminish the Earth in the same way as the loss of the paintings of Van Gogh would diminish our world. It would be the loss of something that enriched the Earth and that cannot be recovered. Unfortunately the preservation of the polar bear requires millions of square kilometers, while a few hectares of museum space suffice for the Van Gogh paintings. In this sense the former is more difficult.

The 'Nature' we find today seems to many of us to be something that has always been since the days of creation. But as the world came out of the last ice age only some 10,000 years ago, natural and anthropogenic changes seem to have proceeded in parallel. The Amazon basin was much drier during the Younger Dryas with the river discharge 40% lower than today; it has been increasing ever since with implications for the biodiversity in the rainforest [43]. As a result, over the last 3,000 years the rainforests of eastern Bolivia have been expanding further south than at any time during the last 50,000 years [44]. So what is the 'natural' range? It was generally thought that central Amazonia was virgin forest, untouched by humans. But when Francisco de Orellana and his small band of adventurers first descended the river, they reported dense concentrations of villages [45] which later were destroyed by European diseases. Yet the widespread fertile 'terra preta', full of charcoal and pottery shards, documents a long human occupation of parts of Amazonia with a sophisticated agriculture [46]. Archeologists had 'known' that there was nothing to be found in Amazonia, and so no one had looked!

In the same way much of the North American forest, including that in natural parks, had actually been reduced earlier during the 19th century when firewood was used for heating and transport. In Australia many natural areas have been shaped by changing climates and by the fires set by the Aborigines. And all around the Arctic the recent warming is changing the tundras and boreal forests even in areas untouched by human hands. The Sahara desert is an ecosystem with species well adapted to drought, but if we had gone there 6,000 years ago we would have found a relatively humid climate, with lakes and ample and varied vegetation and animal life [47]. So, if we preserve pristine conditions, we preserve a snapshot of an ecological system that is continuously changing. What we protect today will in any case be different tomorrow because of natural climate variability. Of course, this is *a fortiori* because with the rapid climate change taking place now such effects are becoming more prominent even in places untouched by humans. A recent study in the Amazon rainforest found that out of 115 abundant tree genera in undisturbed forest plots, 27 had changed significantly in frequency over only 15 years, which may perhaps be ascribed to

increasing CO_2 abundance [48]. So if someone returned there in a century, he might find a very much changed forest. Of course, this is no reason to cut down trees that have become more abundant.

8.3.1 Deforestation

Forests play an essential role in the Earth's ecology. Forest cover mitigates the erosive effects of rain and wind; a large part of plant and animal species live in forests, in particular in the tropics; and the ability of forested areas to hold water for some time reduces the impact of the alternation of droughts and floods. The loss of forests by human activities can therefore have serious consequences, as many societies have experienced to their detriment. Prudent utilization of forests has given many benefits to humans. Wood has been much used in the building of our dwellings, in protecting us against the cold and in the cooking of our food. As long as the logging was a relatively minor perturbation to the forests, regrowth could maintain the forested area without too many problems. On a timescale of less than a century most of the dwellings constructed would rot away, returning their mineral content to the soil, and the ashes of the fires would do so even more rapidly. Also, the CO_2 released in burning wood would enter the atmosphere but would soon be used in photosynthesis and thus would not accumulate. Similarly, the oxygen that is used up in the burning process or in other oxidation processes would be restored by the next generation of plants and trees. This shows the partial fallacy of the statement that is frequently made that the tropical forests are the 'lungs of the planet', the loss of which would deprive us of oxygen. In fact, the present atmospheric oxygen is the result of millions of years of photosynthesis. Even if every plant or tree on Earth was cut down, it would take millions of years before significant oxygen loss would be noticed.

As the world's population increased, significant deforestation ensued. There is not yet full agreement on the precise course of events, but it is generally agreed that already some 3,000–4,000 years ago deforestation around the Mediterranean was extensive. Suggestions that much earlier deforestation had become sufficient to affect climate remain controversial [49]. As time went on, deforestation on the Eurasian continent increased, first in the temperate regions and later extending also into the tropics. In North America much wood was cut, especially in the 19th century, while in much of South America deforestation was increasing towards the end of the 20th. Sometimes deforestation was remarkably rapid. In Madagascar, there was still 65% of the tropical rainforest in 1950, but only 10% in 2000 [50]. Examples of the disastrous effects of deforestation abound [51]. After the forest is cut, rain and wind take away the fertile soil releasing a significant amout of CO_2.

More recently in the temperate regions the situation has been reversed. As agriculture became ever more efficient, and as the population stabilized, agricultural needs could be met with smaller areas. In addition, as coal and later oil and gas replaced firewood, the pressure on forests diminished, while unused agricultural areas regained forest cover. In fact, in Europe it has been estimated that forest cover passed through a minimum of only 5% of the total area around 1700–1800 and has been increasing since to reach some 30% by

AD 2000 [52]. In North America the minimum (35%) was reached around 1900 [52]. In China where deforestation had reached alarming proportions, afforestation projects have stemmed the decline and over the last 25 years planted forests have been increasing [53].

The situation in the tropics is very different. Here rapidly increasing populations and less efficient agriculture are continuing to push back the forests, and the overall forest cover has been reduced from 60% to 40% [52]. Moreover, in some countries, such as Indonesia, intensive logging of tropical hardwoods (mahogany, teak, etc.) leads to much damage even in so-called protected areas that are not fully deforested [54]. The most important areas of contiguous tropical forest are Amazonia (5.3 million km^2), central Africa (2 million km^2) and Borneo (0.5 million km^2). The species' diversity in the tropical rainforests is incredible – with ten 1-km^2 plots in Borneo yielding 700 species, which is as many as there are in all of North America [54]. Probably the most immediate risk to the tropical forests is in Borneo. The Amazon basin, which accounts for some 45% of the world's tropical forest, is still favored by inaccessibility, but in many other areas the situation is rapidly deteriorating. Again, the increasing emphasis on biofuels is likely to have a catastrophic effect. The same climatological circumstances that have caused the tropical forests to be so rich are favorable for the cultivation of plantations of oil palms and other crops that can be converted into fuel. An intensifying deforestation results in several places. In just the last 10 years 60% of the forests in the Rian province of Sumatra, with an area of some 4 million hectares, has been cut to make space for pulpwood and palm oil plantations. This has resulted in incredible destruction and heavy pollution [55].

The humid tropical forest cover and the annual rate of deforestation/degradation are shown in Figure 8.5 [56]. At these rates it would take 200 years for the forest to disappear in Latin America and Africa, but less than 80 years in South-East Asia. There has also been some reforestation amounting to 10–15% of the deforestation/degradation [56]. However, the secondary forest is not all equivalent to the original one.

The Amazon basin (Figure 10.12) has received particular attention because it contains nearly half of the world's tropical forests. As of 2003 the region had 530 million hectares (Mha) of closed-canopy forest – about 85% of the original area [57]. Some 360 Mha are situated in Brazil where rather detailed studies have been made [58]. Early estimates of current deforestation rates appear to have been excessive, but data from the Landsat satellite have allowed more satisfactory results to be obtained. From these and, more recently, from the Spot satellite (see Chapter 10), it is found that the deforested area in Brazil was about 8 Mha in 1978, 23 Mha in 1988, 39 Mha in 1998 and 49 Mha in 2003, corresponding to a slowly increasing rate of some 0.6% per year [59]. Brazilian plans for the construction of several all-weather roads will tend to increase the inflow of agricultural workers. It has been estimated, on the basis of model studies, that if past practices – including the disregard of Brazilian environmental laws – were to continue, by 2050 no more than 320 Mha of forest would remain

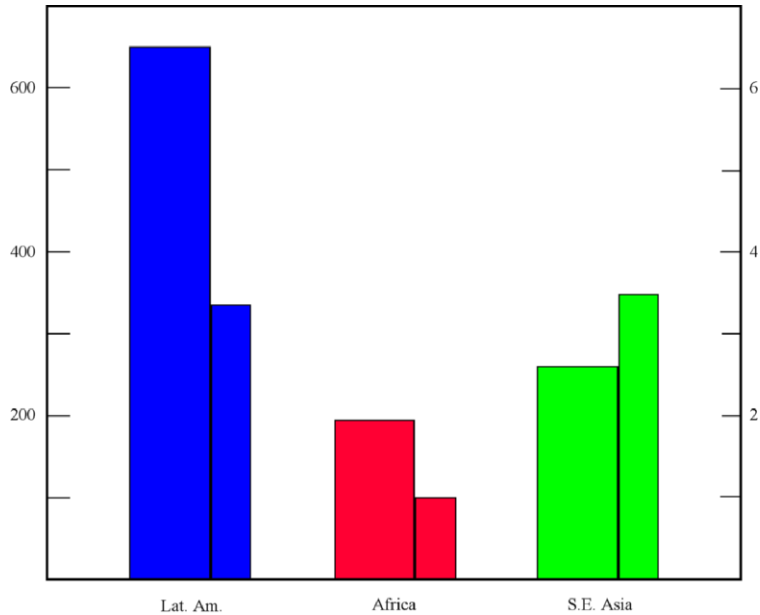


Figure 8.5 Humid tropical forests and deforestation in three areas. The broad bars indicate the forested areas (left scale) and the narrow bars the sum of annual deforestation and degradation, both in millions of hectares.

in the Amazon basin, about half of the original area [57]. However, if the legally protected areas were preserved and if the environmental laws were fully enforced, an additional 130 Mha of forest could be preserved, and greatly improve the situation. Most climatological models suggest that this may be inadequate to maintain the current rainfall in the basin which results in part from transpiration by the trees [60], but there is no unanimity on this [61].

Measurements of deforestation give only a partial impression of the damage [62]. In fact, in 2005 some 1.9 Mha of forest were lost by clear cutting and an additional 1.2 Mha was affected by selective logging, of the most valuable trees, which also impoverishes the forest. Moreover, deforestation leaves much inflammable material on the forest floor, and the resulting forest fires do additional damage to the forest and its denizens. Furthermore, deforestation does not usually begin with the clear cutting of a large area all at once; it is more a patch here and another patch somewhere further away. Thus, the remaining forest may consist of isolated patches or be very close to deforested areas, neither of which is favorable to the preservation of wildlife. In fact, it was estimated that in 1988 the forested area less than 1 km distant from deforested areas was 1.5 times larger than the deforested areas themselves [58].

In addition to forest losses due to logging, climate change could have major effects. Unfortunately in the most important area, Amazonia, different models are producing different results. However, it is generally thought that increased

temperature would lead to a drying of the soil and thereby to a transformation of the rainforest into savannah-like vegetation, at least in the eastern parts. Since a substantial part of the rainfall is secondary and produced by the evaporation from tree leaves, this can lead to rather large shifts. Also savannah contains less carbon and so such a transition would inject a significant amount of CO_2 into the atmosphere. Of course, in areas where the trees have been cut, the effect is even stronger.

8.4 Conclusion

Water will be scarce in many places where there is an increase in population but, in the long term, carefully executed waterworks and desalination should improve the situation. Agriculture should be able to feed the world an adequate vegetarian diet, and probably more than that; however, much care will be needed to preserve the soil and only agriculture with a long-term sustainability perspective will be able to avoid more disastrous possibilities. The preservation of the tropical forests will require an immediate drastic reduction and, in the near future, elimination of the deforestation, although this will require a substantially increased cooperation between developing and developed countries.

8.5 Notes and references

[1] Postel, S.L. et al., 1996, 'Human appropriation of renewable fresh water', *Science* **271**, 785–788.

[2] The values of the fluxes and reservoirs for the hydrological cycle (Figure 8.1) are not very certain. River runoff estimates range from 34,000 to 47,000 km^3 per year with recent values of 40,000 km^3 per year (see references [1] and [3] and 45,500 km^3 (see [4]). Groundwater ranges from some 10 million km^3 (see [1]) to 23 million km^3 (see [4]).

[3] Vörösmarty, C.J. et al., 2000, 'Global water resources: vulnerability from climate change and population growth', *Science* **289**, 284–288.

[4] Oki, T. and Kanae, S., 2006, 'Global hydrological cycles and world water resources', *Science* **313**, 1068–1072.

[5] Clarke, R. and King, J., 2004, *The Atlas of Water*, Earthscan, London, pp. 45, 61.

[6] Stanley, D.J. and Warne, A.G., 1993, 'Nile delta: recent geological evolution and human impact', *Science* **260**, 628–634.

[7] Gong, Gwo-Ching et al., 2006, 'Reduction of primary production and changing of nutrient ratio in the East China Sea: effect of the Three Gorges Dam?', *Geophysical Research Letters* **33**, L07610, 1–4.

[8] Fenwick, A., 2006, 'Waterborne infectious diseases – could they be consigned to history?', *Science* **313**, 1077–1081.

[9] Micklin, P.P., 1988, 'Desiccation of the Aral Sea: a water management disaster in the Soviet Union', *Science* **241**, 1170–1176.
[10] Greenberg, I., 2006, 'A vanished sea reclaims its form in Central Asia', *International Herald Tribune*, April 6.
[11] Clarke, R. and King, J., 2004, *The Atlas of Water*, Earthscan, London, pp. 64–65.
[12] Nordstrom, D.K., 2002, 'Worldwide occurrences of arsenic in ground water', *Science* **296**, 2143–2145. For possible remedies see Ahmed, M.F. et al., 2006, 'Ensuring safe drinking water in Bangladesh', *Science* **314**, 1687–1688.
[13] Clarke, R. and King, J., 2004, *The Atlas of Water*, Part 7, Earthscan, London.
[14] Morin, H., 2006, 'De l'eau de mer dessalée pour abreuver Belle-Ile et ses visiteurs', *Le Monde*, 23/24 Juillet, p. 7.
[15] Service, R.F., 2006, 'Desalination freshens up', *Science* **313**, 1088–1090.
[16] IPCC, 2001, *Third Assessment Report*, p. 598.
[17] IPCC, 2001, *Third Assessment Report*, p. 568.
[18] de Wit, M. and Stankiewicz, J., 2006, 'Changes in surface water supply across Africa with predicted climate changes', *Science* **311**, 1917–1921.
[19] Bradley, R.S. et al., 2006, 'Threats to water supplies in the tropical Andes', *Science* **312**, 1755–1756.
[20] Barnett, T.P. et al., 2005, 'Potential impacts of a warming climate on water availability in snow-dominated regions', *Nature* **438**, 303–308.
[21] Diamond, J., 1997, *Guns, Germs and Steel: The Fates of Human Societies*, Norton, New York.
[22] Malthus, T., 1798, *An Essay on the Principle of Population*, Harmondsworth; Penguin.
[23] Nosengo, N., 2003, 'Fertilized to death', *Nature* **425**, 894–895.
[24] Carson, R., 1962, *Silent Spring*, Houghton Mifflin, Boston.
[25] Tilman, D. et al., 2002, 'Agricultural sustainability and intensive production practices', *Nature* **418**, 671–677.
[26] Estimates based on not always concordant data in references [27], [28] and [29].
[27] Houghton, J.T. et al., 2001, *Climate Change 2001: The Scientific Basis*, Cambridge University Press, p. 192.
[28] Morris, D.W., 1995, 'Earth's peeling veneer of life', *Nature* **373**, 25.
[29] The early 2006 FAO database.
[30] Myrdal, G., 1967, *Asian Drama*, Random House, New York (3 volumes), pp. 1396–1397.
[31] Myrdal, G., 1967, *Asian Drama*, Random House, New York (3 volumes), chapters 10.5 and 22.4.
[32] Waggoner, P.E., 1996, 'How much of the land can be spared for Nature?', *Daedalus* **125** (3), 73–93.
[33] Smil, V., 1997, 'Global population and the nitrogen cycle', *Scientific American* **277** (1), 76–81.

[34] Emsley, J., 2001, *Nature's Building Blocks*, Oxford University Press, p. 124.
[35] Lal, R., 2004, 'Soil carbon sequestration impacts on global climate change and food security', *Science* **304**, 1623–1627.
[36] Pimentel, D. et al., 1995, 'Environmental costs of soil erosion and conservation benefits', *Science* **267**, 1117–1122.
[37] Liu, J. and Diamond, J., 2005, 'China's environment in a globalizing world', *Nature* **435**, 1179–1186.
[38] Huang, J. et al., 2002, 'Enhancing the crops to feed the poor', *Nature* **418**, 678–684.
[39] Normile, D., 2006, 'Consortium aims to supercharge rice photosynthesis', *Science* **313**, 423.
[40] Sanchez, P.A. et al., 1982, 'Amazon basin soils: Management for continuous crop production', *Science* **216**, 821–827.
[41] See reference [1].
[42] Dugger, C.W., 2007, 'By disregarding Western advice, Malawi becomes a breadbasket', *International Herald Tribune*, December 3, p. 7.
[43] Maslin, M.A. and Burns, S.J., 2000, 'Reconstruction of the Amazon basin effective moisture availability over the past 14,000 years', *Science* **290**, 2285–2287.
[44] Mayle, F.E. et al., 2000, 'Millennial-scale dynamics of southern Amazonian rain forests', *Science* **290**, 2291–2293.
[45] Prescott, W.H., 1847, *The Conquest of Peru*, Book IV, chapter IV, Capitulacion con Orellana.
[46] First International Workshop on anthropogenic terra preta soils, Manaus, Brazil, 13–19 July 2002.
[47] See Chapter 6, note [13].
[48] Laurance, W.F. et al., 2004, 'Pervasive alterations of tree communities in undisturbed Amazonian forests', *Nature* **428**, 171–174.
[49] Ruddiman, W.F., 2005, *Plows, Plagues and Petroleum*, Princeton University Press.
[50] de Wit, M.J., 2003, 'Madagascar: heads it's a continent, tails it's an island', *Annual Review Earth and Planetary Sciences* **31**, 213–248.
[51] Diamond, J., 2005, *Collapse*, Penguin Books, London.
[52] IPCC, 2001, *Third Assessment Report*, WGII, p. 310.
[53] Fang, J. et al., 2001, 'Changes in forest biomass carbon storage in China between 1949 and 1998', *Science* **292**, 2320–2322.
[54] Curran, L.M., 2004, 'Lowland forest loss in protected areas of Indonesian Borneo', *Science* **303**, 1000–1003.
[55] Gelling, P., 2007, 'For pulped forests a survival trade-off', *International Herald Tribune*, 6 December, p. 6.
[56] Achard, F. et al., 2002, 'Determination of deforestation rates of the world's humid tropical forests', *Science* **297**, 999–1002.
[57] Soares-Filho, B.S. et al., 2006, 'Modelling conservation in the Amazon basin', *Nature* **440**, 520–523.
[58] Skole, D. and Tucker, C., 1993, 'Tropical deforestation and habitat

fragmentation in the Amazon: satellite data from 1978–1988', *Science* **260**, 1905–1910; updated to 2003 (in reference [59]).
[59] Laurance, W.F. et al., 2004, 'Deforestation in Amazonia', *Science* **304**, 1109.
[60] Silva Dias, M.A. et al., 2002, 'Cloud and rain processes in biosphere-atmosphere context in the Amazon region', *Journal of Geophysical Research* **107**, 8072.
[61] IPCC, 2001, *Third Assessment Report*, p. 443.
[62] Nepstad, D.C. et al., 1999, 'Large-scale impoverishment of Amazonian forests by logging and fire', *Nature* **398**, 505–508.

9 Leaving Earth: From Dreams to Reality?

We can do anything we want. We can say anything we want to ourselves, because it is easy to fool ourselves. But, we cannot fool Nature. And if we try to fool Nature, we only court disaster.

Richard P. Feynman

9.1 Introduction

With the perspective of an increasing population, with the limits faced on the most crucial resources required for sustaining life, with the deterioration and the warming of the climate, plus the possibilities offered by space techniques, the option of leaving a planet that may become inhabitable if not properly managed is an alternative to our future that many have envisioned. Finding hospitable outposts outside the Earth through the Solar System has in fact very often been addressed. Even without any pressing needs, humans have for a long time dreamt of escaping the Earth and exploring the Universe around it. At the turn of the 17th century, in his *Somnium seu Astronomia Lunari – A Dream or Astronomy of the Moon,* with a lot of fantasy but also accuracy, Kepler described how the Sun and its planets would appear to an inhabitant of the Moon. He imagined the living creatures of that new world where the length of the day was different, the temperatures were different and the seasons were different, but where the laws of celestial mechanics would be the same as on Earth.

Kepler's *Somnium is* a perfect blend of fantasy and scientific rigor, one of the first serious science fiction books. The dream was later echoed by the Russian school teacher Konstantin Tsiolkovsky who, in the late 19th century, claimed that '*The Earth is the cradle of the mind, but we cannot live forever in a cradle'.* The development of the first intercontinental rockets, the launch of the first artificial satellite and of the first cosmonaut by the Soviets in 1957 and 1961, offered an illustration that this dream might one day be realized. The first landing of humans on the Moon in 1969 by the Americans, for the first time in the history of humanity, confirmed the prediction that humans might live outside the environment in which they were born have developed and evolved. Since then, terrestrial robots have extensively traveled through the Solar System, landed on the Moon, Venus, Mars, Titan, asteroids and soon on comets. They have reached the limits of the heliosphere and are starting a long journey in interstellar space. For a very long time the popular topic of many fiction books, the dream of

exploring interstellar space has been realized. Extending the realm of civilization to places other than the surface of the Earth can now seriously be discussed. However, at just one light-second to the Earth, the Moon is still the most distant outpost where humans have physically traveled. Is it conceivable that we could go further out, or is it just another of these utopias that will feed science fiction books and haunt the imagination of astronauts and some famous scientists? In this chapter we discuss the realism of such an alternative to our future. But where should we go? How shall we get there, and for how long?

9.2 Where to go?

Our goal here is to analyze where, and discuss how, to settle in the next centuries for 100,000 years or so, until that hypothetic new habitat would in its turn become inhospitable. In the course of that exercise, we should compare the living conditions offered thereon with those existing or foreseen on Earth in the future. As yet, the Moon is the only place in the Solar System on which humans have landed. With a radius of only 1,738 km it is a relatively small body and cannot permanently accommodate a large share of the Earth's population. However, it has been demonstrated by Apollo that our satellite is definitely reachable, and it may soon be revisited or inhabited permanently. At the end of this chapter we discuss that option and analyze what to do with the Moon in a context where we continue to live on Earth. We also further discuss the possibilities of finding Earth-like planets in orbit around other stars as well as the difficulties of reaching them in the present state of the technology.

This leaves us with very few options, as we consider only planets – or moons of planets – which are sufficiently large and possess physical characteristics such that they could host a substantial portion of the present Earth's population. Within this restricted set we find Venus, Mars and Titan. All three possess an atmosphere, even though none is presently breathable for a normal human being. All three are inhospitable and would require a substantial amount of transformation, or 'terraforming' [1]. Table 9.1 presents the main physical characteristics of these three bodies as compared to the Earth and Figure 9.1 shows images of all three and the Earth, as obtained from space.

The Earth, Venus and Mars have followed different evolution scenarios because of their distance to the Sun and of their different physical characteristics. All three have CO_2 in their atmosphere: nearly 97% for Venus, 95% for Mars and 380 ppm for the Earth. This difference is explained by the absence of tectonics on both Venus and Mars, making it impossible to recycle CO_2 between their crust and their atmosphere [2]. In the case of Venus, this resulted in a runaway greenhouse situation which led to a complete loss of water, with surface temperatures unsuitable for any kind of life. In the case of Mars, its smaller physical dimension has played a key role in its evolution as it was not able to retain its atmosphere, and what remains of it today is insufficient for maintaining a high enough temperature and liquid water, making Mars a frozen

Table 9.1 Main physical characteristics of Venus, Mars and Titan as compared to the Earth

Planets characteristics	Venus	Earth	Mars	Titan
Distance from Sun (AU)	0.72	1.0	1.52	10
Radius (km)	6,052	6,376	3,403	2,575
Surface pressure (bars)	92	1	0.006	1.5
Solar constant at the planet(Wm^{-2})	2,620	1,366	594	14
Mean surface temperature (K)	730	288	210	94
Length of day (Earth days)	243	1	1.025	16
Radiative equilibrium or effective temperature T_e K)	230	255	212	85
Albedo [3]	0.76	0.367	0.15	0.2
Main atmospheric composition (%)	N_2: 3.5 CO_2: 96.5 Ar: 0.007	N_2: 78.1 CO_2: 0.0382 O_2: 20.9 H_2O: 0 to 4 Ar: 0.93	N_2: 2.7 CO_2: 95.0 O_2: 0.13 H_2O: 0.03 Ar: 1.6	N_2: 95 CH_4:1 to 5 Ar: 3.3 10^{-5}

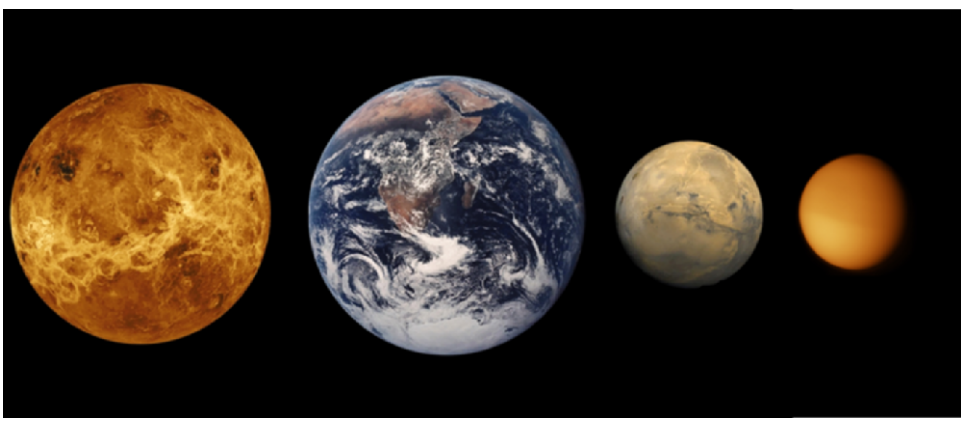


Figure 9.1 Size comparison of possible habitable planets (left to right): Venus, Earth, Mars and Titan. The Earth and Venus are nearly the same size while Mars has a diameter half that of the Earth and Titan a little smaller. The Venus image is a radar image from NASA's Magellan mission which could observe through the clouds of the thick atmosphere of the planet. (*Source*: NASA–JPL.)

desert. This trend in the evolution of both Venus and Mars ought to be kept in mind when it is envisaged to return them to a situation suitable for maintaining life.

The fourth power of the effective temperature of a planet T_e (see Chapter 2) is directly proportional to the total solar irradiance – or solar constant – at the planet's orbit [3]. Models of solar evolution show that the solar constant varies with time at a rate of 1% per 100 million years. As evaporation of gases and water

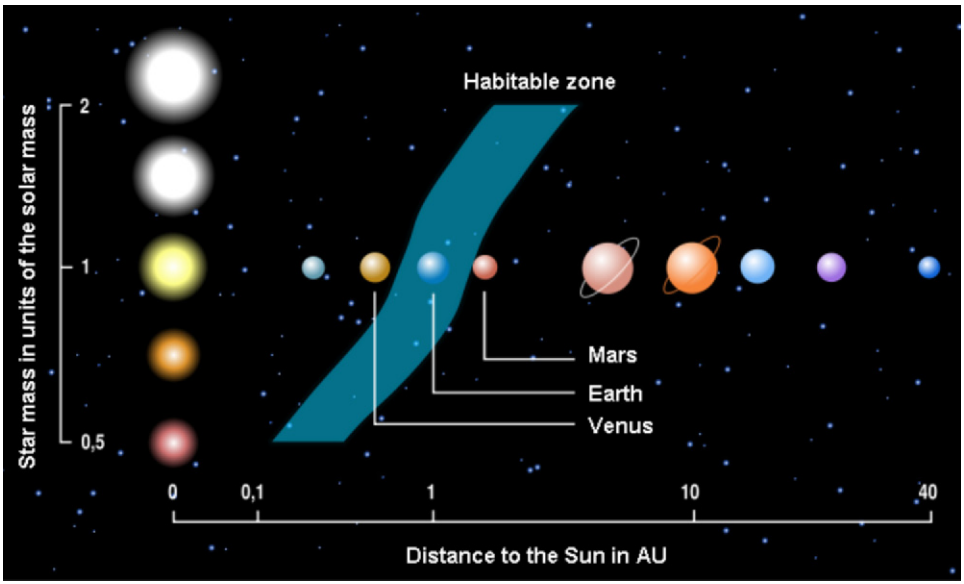


Figure 9.2 The habitable zone of the Solar System where conditions exist for water to be liquid on the planet's surface. The Earth presently occupies the center of the zone.

vapor proceeds, the albedo [3] may also vary, depending on the cloudiness or ice coverage of the planet's surface. These two phenomena influence the effective temperature and therefore the evolution of the climate. Surprisingly, for Venus and the Earth, whose surface temperatures are definitely well above the freezing point of water, T_e is below. The presence of a greenhouse effect is therefore necessary to explain this difference. As shown on Figure 9.2, the combination of the greenhouse effect induced by the constituents of its atmosphere (CO_2, and H_2O at a pressure of 1 bar) and its distance to the Sun, place the Earth in the middle of the habitable zone, making it possible for water to remain liquid at the surface. On the contrary, both Venus and Mars sit very close to the extreme borders of the habitable zone and are just missing the appropriate conditions that allow water to remain liquid. If it were to be decided to implant future outposts or colonies on either Venus or Mars, the peculiar positions of these two planets require some mammoth engineering to correct for the drawback that the Earth has fortunately avoided. We shall now analyze the two cases separately.

9.2.1 The case of Venus

Why has Venus become so hot and so dry?

With an albedo of 0.76, Venus should be nearly 60°C colder than the Earth today. As said, its very high surface temperature of 730 K, ($\sim$500°C higher than the effective temperature), is due to the enormous atmospheric pressure and to the extreme greenhouse effect produced early in its life by the high concentra-

tion of CO_2, re-enforced by the presence of water vapor (now nearly totally absent) and cloud particles made mainly of droplets of sulfuric acid produced by SO_2 of volcanic origin. At this temperature, life is impossible. The Russian probes that landed on Venus in 1970 and 1982 could transmit signals and pictures of the surface for no more than 2 hours maximum (Venera 13), after which the equipment was destroyed by the heat. This makes the implantation of life on the planet a genuine technological and biological challenge. If we assume, as is usually done, that both the Earth and Venus – whose physical dimensions are nearly identical – had the same initial conditions at the time of their formation, their evolution has clearly been drastically different. Re-establishing these initial conditions on Venus through terraforming would require that we understand properly what happened to the planet early in its life. This has been and is still the subject of intense discussions [2, 4, 5], and we will just summarize our present – and not definite – understanding of what determined the evolution of Venus from the time of its formation to the situation where it stands today: an extremely hot and dry planet with an unbreathable atmosphere more than 90 times the thickness of our own. New scenarios for the evolution of the planet may be elaborated when new data become available.

It can be assumed that originally Venus had an amount of water equivalent to that of the Earth [2]. Due to its proximity to the Sun – possibly because of a lower albedo than at present – and the originally modest but regularly growing greenhouse effect of water vapor and CO_2, the surface temperature of Venus was gradually raised to the extent that water could not remain liquid. When the surface temperature reached 70°C, the lower atmosphere contained 20% of water vapor per volume. Normally, as on Earth, water would be trapped in the cold region of the atmosphere at the tropopause level situated at about 8 to 9 km between the troposphere and the stratosphere (see Figure 10.4). However, the higher temperature of Venus – plus the warming resulting from the condensation of water which releases heat – would result in a slower cooling of the atmosphere, pushing the 'cold trap' at higher and higher altitudes to the point where, above 50–60 km, ultraviolet radiation would break H_2O into its two constituents. The relatively light hydrogen atoms would escape into the interplanetary medium, thereby depleting the water content of the planet.

This scenario is consistent with the fainter luminosity of the young Sun (see Chapter 2, Figure 2.7) which would provide a solar constant at Venus of 1.4 times the present Earth value (instead of 1.9 now), because it requires values between 1.1 and 1.4 to be triggered. It is also compatible with the high deuterium/hydrogen ratio which for Venus is 100–150 times the value on Earth as measured by the NASA Pioneer mission in 1978 and more recently by ESA's Venus Express mission [6], the heavier deuterium atoms escaping less easily from Venus than hydrogen. Without water, CO_2 could not be recycled in the form of carbonates in the soil as is the case on Earth. Being too heavy to escape into outer space, it could only reside in the atmosphere. In effect, its present amount has been estimated to be equal to that stored as carbonates in the Earth's soil. Furthermore, the absence of water, which plays an essential lubrication role in

Earth's tectonic activity, has probably put a stop to any such activity that might have existed on Venus, now a single-plate planet.

What would happen to the Earth if its surface temperature were raised, as has been the case on Venus? The average surface temperature that would turn the Earth into a Venus situation and create a similar CO_2 atmosphere is around 340 K, that is 52°C above the present value. These estimates ought to be considered as rough figures because the effects of the clouds on the albedo must be evaluated in more detail. The 'cold trap' mentioned earlier would be gradually raised to higher altitudes as the surface warms. When the surface temperature reaches values above 70°C, a Venus situation is triggered and water starts to be photodissociated. It has been estimated that in a model where the effect of clouds on the albedo is not taken into account – and with a solar flux 1.1 times the solar constant at Earth – our planet, due to its lower albedo than Venus, would lose its water content through the same photodissociation process which dried Venus nearly completely [2, 7]. This means that if the Earth were 5% closer to the Sun at 0.95 AU (where that condition would be fulfilled), life would probably never have developed. Since solar luminosity is increasing by 1% every 100 million years, the time left for Earth to retain its water and life is about 1 billion years.

Terraforming Venus

Quite obviously, in order to render Venus habitable, one must first of all solve the main problem that made it uninhabitable – that is, re-hydrate it, remove the 92 bars of CO_2 from its atmosphere and cool it down. The first proposal to 'engineer' Venus was by Carl Sagan in the 1960s. His original ideas were followed by several others [8]. One scenario to get rid of CO_2 is to fix the carbon through the import of biological species. To recreate a humid environment it would be necessary to import the hydrogen which was lost in the early days, and the giant planets, in particular Uranus, seem to be the most interesting sources for the light atom! The time estimated to deliver the required amount of hydrogen to Venus is 15,000 years. This represents a substantial fraction of the 100,000 years we are dealing with here. To achieve that, it would be necessary to have a minimum fleet of 150 independent ferries, in continuous use, traveling on the Uranus–Venus transfer orbit (period $\sim$31 years) and arriving at Venus every 226.1 days, which is the minimum energy transfer orbit from Venus to Uranus. This has been considered by Freeman Dyson, a professor at Princeton University, to be too much of a conservative estimate, which he lowered to close to 500 years [9]. The factor of 30 difference between these two estimates is enormous and illustrates the difficulties of conducting such scenarios. The ferries would carry hydrogen containers of roughly 10 km across! The requisite amount of iron to build the containers would come from the mining of Uranus moons or of asteroids, providing we move them to the orbit of Uranus! Practically, it would in any case require the development of new powerful cargo rockets probably powered by engines that do not yet exist and are never likely to exist!

In parallel, solar illumination would have to be decreased to avoid running

into the same thermal problem again, and several options have been considered. One, proposal [9] is to shield Venus by a sunshade that would decrease the solar constant by a factor of 2, bringing it to the value at the Earth orbit. The complete picture to 'terraform' Venus should also consider reducing the solar day to make it as close as possible to 24 hours instead of 243 times that value in order to bring the Venus climate closer to what it is on Earth. That could be achieved by spinning up the planet, and several solutions have been considered, such as magnetic torquing [10]. The net result of that gigantic, very imaginative, engineering process is that it is highly improbable, not least because there is no way to properly evaluate the feedback mechanisms that, on Earth, are sources of the present concerns on climate. The final product would probably be far from the ideal conditions one hoped to re-establish, and it appears that the Earth itself, as imperfect as it is, would most likely offer a better living than this new world whose environment would have to be monitored much more closely and continuously than that on Earth at present.

Terraforming Venus would seem to be more a science fiction concept – or at best an undergraduate student intellectual exercise – than a serious scientific and engineering scenario, even though, with more exclusive dreaming, one could always consider that future scientific and technical progress might solve the present impossibilities. Furthermore, as is well illustrated in the relevant literature, such a utopian undertaking would take not only centuries if not millennia to be completed, but would also necessitate gigantic financial resources and, not the least of problems, multi-century political commitments. This is unthinkable today! But perhaps in the future? Who can tell? However, the decision must be taken now to start the long process! The assumption that governments or even a future world governance, as discussed in Chapter 11, would agree to spend their resources on such an undertaking hundreds or thousands of years before completion without being assured of its final success, is definitely not credible. And we have not even addressed the most profound issue of irreversibly destroying a natural environment, without being assured that this will solve our present concerns – certainly losing a precious reference in the history of the natural evolution of the Solar System. Let us quote the father of Venus terraforming, C. Sagan [11]: '*All proposals for terraforming Venus are still brute-force, inelegant and absurdly expensive. The desired planetary metamorphosis may be beyond our reach for a very long time, even if we thought it was desirable and responsible.*' At this point, there is no need to consider further this scenario if even its own father repudiates it! It nevertheless reveals the high sense of utopia that can animate the thinking of scientists who usually have a reputation for being extremely serious. We should then allow Venus, with its infernal greenhouse effect, to continue to serve us as the best example of what we may expect on Earth if we don't manage our own habitat properly. We now turn ourselves to Mars, the other potential candidate for terraforming.

9.2.2 The case of Mars

Why has Mars become so cold and dry?

Mars is considered to be – after the Earth – the most hospitable planet in the Solar System and many essays have been published in the period following the landing of NASA's Viking mission in 1976, on the terraforming or ecopoeisis [12] of the red planet, which has received more serious attention than Venus [13]. The various scenarios envisaged rest on the fact that Mars has nearly all the ingredients that would make the planet habitable. Its larger distance to the Sun and its poor thermal inertia associated with a dry and sandy soil, make its temperature very cold, at about minus 60°C on average, with, however, pleasant maxima of 20°C in the austral summer mid-days, but unfortunately followed by minima of minus 100°C during the night, and some minus 120°C at the poles. Human beings might adapt to such temperatures. However, the atmosphere of Mars, as that of Venus, made essentially of CO_2 with a pressure of 0.006 bar, is too thin to keep water in the liquid phase that is necessary for life. Its greenhouse effect is just able to raise the temperature by about 6°C, and cannot offer liveable climate conditions. In addition, the thin atmosphere lets the lethal ultraviolet radiation from the Sun pass through, killing any form of life that might exist on the surface. The result is what we see today: a dry cold planet, with almost no atmosphere, on which life is probably impossible to sustain except, perhaps, underground where solar ultraviolet radiation cannot penetrate.

The evolution of Mars has been substantially different from that of Earth because of its relatively small size. Mars, which probably had a liquid core that was sufficiently hot to be able to activate an internal dynamo and generate a magnetic field, cooled down faster than the Earth and plate tectonics could not act to recycle the carbon, as on Earth. The magnetic field weakened and disappeared quite rapidly in some 500 million years within the gradually cooling core. Consequently, the dynamo came to a stop. In the absence of a magnetic field and of a protective magnetosphere, the original atmosphere of CO_2 disappeared through erosion by the solar wind – probably much stronger for the young Sun – which could easily dispose of the 1–2 bars of atmospheric CO_2 [14]. The oxygen atoms, accelerated by the solar wind, would collide with CO_2 molecules that were not tied strongly to the planet because of its low gravity, and eject them into space. In addition, the smaller gravity of Mars allowed the atmospheric water to evaporate more easily than on Earth. Impacting bodies most probably contributed also to the erosion process in the first billion years, as can be inferred from the high craterization of the ancient terrains of Mars [5]. The greenhouse effect gradually became inefficient because its main constituents disappeared as well as the entire atmosphere.

Because its atmosphere is so thin, contrary to Venus which is surrounded by a heavy haze, it is possible to observe in detail the whole surface of Mars and trace back its evolution since nearly the beginning of the formation of the Solar System. The many pictures taken by the American and European missions clearly evidence the existence of many craters and show the presence of dry river

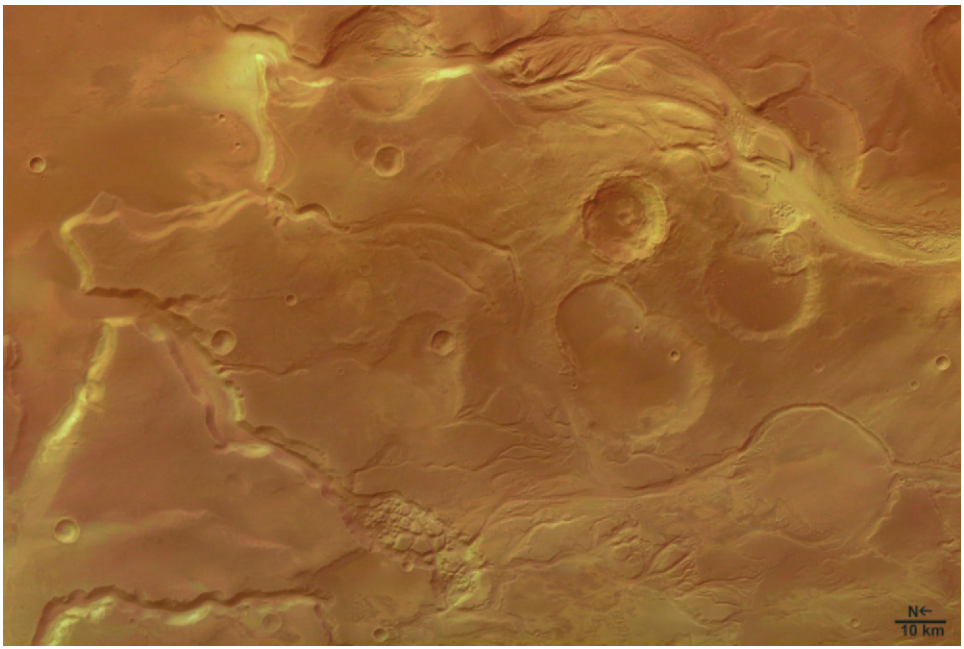


Figure 9.3 This image of fluvial surface features at Mangala Valles on Mars was obtained by the High Resolution Stereo Camera (HRSC) on board Mars Express with a resolution of 28 meters per pixel. The picture evidences a superposition of valleys and craters which allows a dating of the fluvial network. (*Credit*: ESA.)

channels dating from the first 500 million years, revealing a rather wet climate in the early stages of Mars (Figure 9.3). Liquid water, most probably originating from the planetesimals that built the planet [15], existed on Mars but only during the period ending some time between 3.8 and 3.5 billion years ago which corresponds to the Late Heavy Bombardment, in the so-called Noachian eon. This is estimated from the numbers of craters that overlie the valley networks. An early atmosphere of between one and five times the Earth pressure would have been able to keep the water liquid [2]. In fact, it is possible that the original amount of water might have been larger than the amount on Earth. The existence of that originally wet climate has also been confirmed by the recent identification of clays made by Mars Express and found to be existing only in the oldest terrains, indicating that a major climatic change occurred on Mars around 3.5–3.8 billion years ago [16].

Most of the water has disappeared from the surface but there still exist unknown quantities at the poles as permafrost in the underground and even in the form of ice in the middle of craters (Figure 9.4). However, fluvial erosion is rather modest and ancient, making Mars, together with our own Moon, one of the best geological records of the early history of the Solar System. The presence of liquid water in the ancient past testifies to the existence of a sufficiently thick

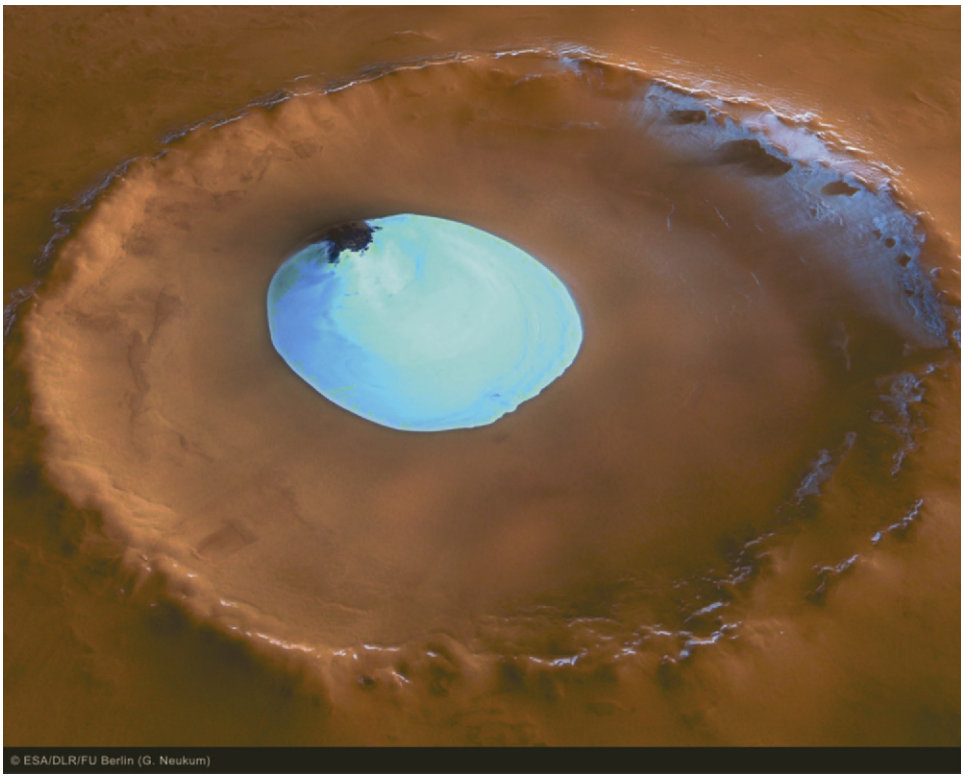


Figure 9.4 Water-ice in the Vastitas crater as observed by the High Resolution Stereoscopic Camera on Mars Express. (*Credit*: ESA.)

atmosphere inducing, as on Earth, a greenhouse effect most probably caused by a mixture of CO_2 and H_2O. However, owing to the larger Sun–Mars distance contrary to Venus, this greenhouse effect did not result in extreme temperatures as these two gases were not abundant enough to raise the temperature above 0°C. The presence of carbonic ice clouds might have been able to trap the thermal emission from the surface, thereby contributing a little more to heating the planet [15]. As on Earth, CO_2 might have dissolved in whatever liquid water reservoirs existed there, and ended up in the form of carbonates in the soil, with the difference that the absence of plate tectonics required a different type of recycling mechanism, probably involving lava from volcanoes. But where are these carbonates? Until now no evidence of their presence has been found. It has been suggested that they may be layered underneath the polar caps [17]. Until it is possible to search the underground of Mars and discover the missing carbon, it will not be possible to understand how CO_2 escaped the red planet, and deprived it from having the greenhouse effect that might have permitted life to develop.

This scenario is far from definite and probably too simple, taking into consideration in particular the cooling by CO_2 clouds due to their high albedo

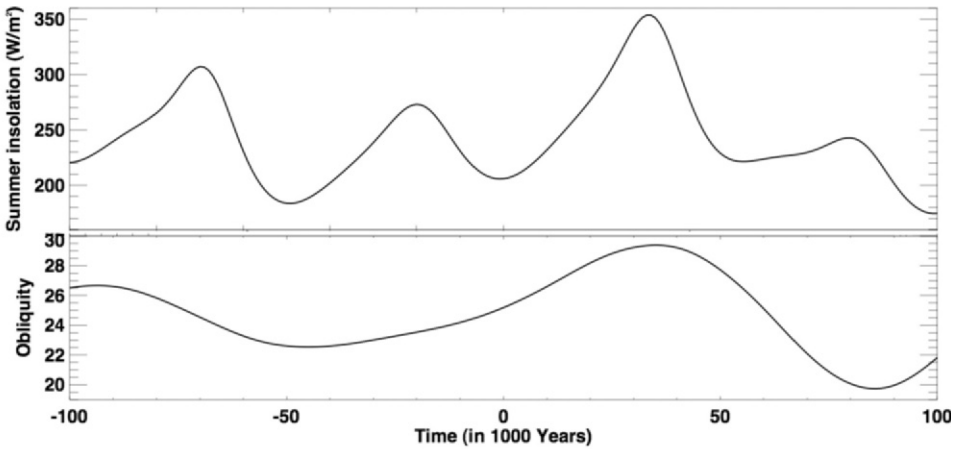


Figure 9.5 Obliquity and insulation at the North Pole surface of Mars at the summer equinox over the last 100,000 years, and for the next 100,000 years. (*Credit*: Laskar [18].)

and the countering warming effect mentioned above. Also, an extra greenhouse effect involving gases other than only CO_2 and H_2O – such as methane or ammonia – should be taken into account [4]. The characteristics of the orbital cycles of Mars induce very strong variations in the Martian climate. The high eccentricity of its orbit (0.093 as compared to 0.017 for the Earth) allows Mars to receive 44% more solar energy at perihelion during the 'austral' Martian summer**.** Summer in the southern hemisphere is 24 days shorter than in the northern hemisphere, with temperatures differences of some 30°C between the two hemispheres. This explains why the southern polar cap disappears nearly completely during the summer, an effect not observed for the northern polar cap. Similarly, winters in the southern hemisphere last longer than those in the north. Furthermore, as shown on Figure 9.5, the Martian climate is unstable due to the chaotic changes in the obliquity of the planet, resulting in an exchange of water-ice between the poles and the equator when the inclination of the polar axis is changing [18, 19]. These climate variations are very much stronger than those observed on Earth, where the Moon has a strong stabilizing effect on the obliquity and on the climate, making it much easier to develop the right conditions for life. It may well explain the periodic occurrence of short, warmer periods with increased greenhouse effects (possibly enhanced by volcanic activity). These large variations ought to be taken into consideration in the terraforming of the red planet to ultimately offer sustainable living conditions.

Is there life on Mars?

Whether life has ever existed on Mars is still a matter of debate. Conditions for its occurrence were certainly not as favorable as on Earth. If life developed, it was most likely in the Noachian eon and it has remained at the stage of single-cell organisms, probably clustered in ice-covered lakes or in underground reservoirs,

with more difficulties to gradually develop on a (Martian) global scale. This represents a genuine challenge for future human visits to Mars and even more so for its terraforming. Indeed, before it is seriously envisaged to colonize the planet it is essential to know whether life is still present on the planet today, either in a pristine primordial elementary state or in a more advanced one, and whether it presents a danger for possible visitors. But how can we detect life if it is not globally spread? The non-detection of life in one spot on Mars would not necessarily mean that it does not exist in another a few kilometers distant. Before sending colonies, or terraforming the planet, one must be absolutely sure that we have explored all possible niches where life might still exist or that we have all possible elements to conclude that Martian life, in whatever form it has eventually evolved, is definitely extinct.

Without doubt, the discovery of past or present Martian life will be a major revolutionary scientific event of profound consequences. If life is found to still exist on Mars, it will be a delicate decision to land humans there. In all cases, future human visits to the red planet should satisfy all the necessary responsible protective measures to avoid erasure or irreversibe alteration of traces that life may have existed anywhere other than on the Earth!

Terraforming Mars

If it is seriously envisioned to render Mars habitable through ecopoiesis and terraforming, four principal modifications should be applied to the environment [20]:

1. The mean global surface temperature should be increased by at least $\sim$60 K.
2. The mass of the atmosphere should also be substantially increased, ideally by a factor of 100 or more, as well as its oxygen and nitrogen fractions.
3. Liquid water must be made available.
4. The surface UV and cosmic-ray flux must be substantially reduced.

In addition, it is necessary to make sure that the stability of the environment is maintained over a sufficient amount of time. In that respect, the variations of eccentricity and obliquity (Figure 9.5) do present a genuine challenge for any planetary engineering of Mars.

Responding to the first two modifications – the other two would normally follow – requires re-creating an atmosphere with a sufficient greenhouse effect. This can be achieved through the degassing of CO_2 from the regolith – the layer of loose, heterogeneous dust, soil, and broken rock covering any planetary body – if it is found to be there. It can be done also through the importation of artificial greenhouse gases – in a way doing to Mars what we are trying to avoid on Earth [21]! Obviously, ploughing Mars to extract all the carbon from the soil is a disproportionate project. The possibility of alternatively feeding the atmosphere with chlorofluorocarbon compounds has also been considered [13]. While a concentration of $\sim$10 ppm of such absorbers would be capable of warming Mars by about +30°C, the absence of an ozone layer would result in the destruction of these compounds in a few hours, so this does not appear to be a

proper solution! Other greenhouse gases might work but they still have to be found. It has also been envisaged to use ammonia produced by the biological engineering of microorganisms, just to conclude that it is certainly very expensive and environmentally destructive [11].

Another way to warm Mars would be to directly heat it with solar energy reflected by large orbiting mirrors [22]. A mirror of 125 km diameter stationed 214,000 km behind Mars could illuminate the South Pole with an additional $\sim$27 TW. This, in principle, should be sufficient to raise the polar temperature by $\sim$+5°C and to evaporate the polar cap. A first estimate of the size of such a mirror would demand about 200,000 tons of aluminum, probably extracted from the Moon or asteroids as discussed further below! The time estimated to complete such grandiose engineering processes is again definitely out of the scale of our present exercise as it might take 100,000 years to first warm the planet and another 100,000 years to modify the atmosphere through planetary scale biology processes [13]. Finally, the astronomical perturbations of the climate would require permanently active countermeasures. The new Mars that would result from all that engineering would certainly represent an extremely complex system implying feedback mechanisms that are unpredictable and probably uncontrollable.

Independently of the fundamental ethical question of our right or duties to destroy or preserve one of the most important relics of the history of the early ages of the Solar System, we come to the same unavoidable conclusion as in the case of Venus: the transformation of Mars, to render it a habitable planet, is not a realistic solution to envisage for the future of humanity. Possibly, at best, we might make Mars suitable for plants, but at what cost [13]? We might also envisage establishing a pressurized underground colony, but is that more pleasant than living on Earth even in deteriorated conditions (see Section 9.2.5)? Even though we might for a moment assume that at some time in the future we may possess the right technologies to do that, we will not abandon the Earth. The population of the Earth will most likely not be in a worse state than the new Martian population. And who will decide who leaves Earth and who stays? If we could attain that level of technological development, we would also have all the means to control the demographic, technical and industrial development on Earth and sustain a liveable environment. This does not imply that we won't live on Mars or on other objects of the Solar System, but that will be for the same reasons that we inhabit Antarctica today: for science, resource exploitation or tourism! We will do it 'attached' to the Earth, 'stuck' to it and not independently or autonomously. We need the mother planet, to where we will continue to return after our space trips to the red planet. Another visionary idea issued by another professor at Princeton University would be to create between the Earth and Mars, a two-planet human civilization, thereby multiplying the chances of survival of humanity in case of a global catastrophe leading to life extinction on any one of them [23]. This could as much as double our long-term survival prospects and would secure the durability of the human genome. This is an interesting concept and a new mine of science fiction stories!

9.2.3 Other worlds

The next body on the list of those that might be suitable for establishing large sets of populations is Titan, well outside the Habitable Zone. Table 9.1 indicates a surface temperature on Titan well below the freezing point of water, equal to –180°C due to its distance to the Sun, and a non-breathable atmosphere 1.5 times heavier than ours, made of 95% nitrogen and a few traces of methane. These numbers unambiguously show the difficulties to make Titan habitable. The reasons that led us to disregard Venus and Mars would be even more applicable in the case of the biggest moon of Saturn. Consequently, we will no longer consider it in the framework of our '100,000 years exercise'.

Europa, the icy moon of Jupiter, is another possibility (Figure 9.6). Located 780 million kilometers from the Sun, its surface is, at –145°C, also too cold to support life as we know it. Its diameter is 3,122 km, nearly that of the Moon (3,476 km). It has an 'atmosphere' mostly made of oxygen of non-biological origin with a pressure 100 billionths that of the Earth! Its surface is exposed to sunlight and is impacted by meteorites and by dust and charged particles trapped within Jupiter's intense magnetic field. All combined, these processes cause water-ice to sublimate, producing hydrogen, which escapes into outer space, and traces of oxygen which are found in the atmosphere. Creating an atmosphere thick enough to maintain a greenhouse effect compatible with life on a body tidally locked to Jupiter – the effect of the large gravity field of the giant planet – would make the terraforming of Europa a totally unreasonable enterprise. Even more so if, when space missions start drilling the icy crust of Europa, it is found that within the suspected underground water-ocean, some forms of life might have developed that, as in the case of Mars, would deserve to be preserved. In these conditions, as for Titan, we do not consider Europa as a realistic option. What's left then are some of the newly discovered planets orbiting other stars.

There are a good 100 billion stars in our galaxy alone. It is believed that most of them have planets and that a small proportion of these might be suitable to host life in our neighborhood, either pristine or 'imported' [24]. Because all planets are much smaller than their parent star, the light they give off is very faint compared to their star and finding them is extremely difficult unless they are huge. Indeed, the techniques used so far (mostly through gravity perturbations induced by the planet on its parent star, which is seen moving slightly in the sky) bias the present set of observations towards giant Jupiter-size objects that are made of gas rather than solid material, are orbiting at close distances to their parent star and completing their orbital years in just a few days, and are far too hot or far too cold for life to survive. However, the continuous refinement of the observational techniques leads to the discovery of more and more of the smaller objects of dimensions approaching those of the Earth.

In that respect, there is hope that space technologies will drastically change the landscape. Several missions planned in the near future in Europe as well as in the United States offer interesting prospects. Also, model simulations promise to be very useful for identifying Earth-like planets. About 5% of the known giant planets may have Earth-like planets [25]. For example, the star 55 Cancri, at 41

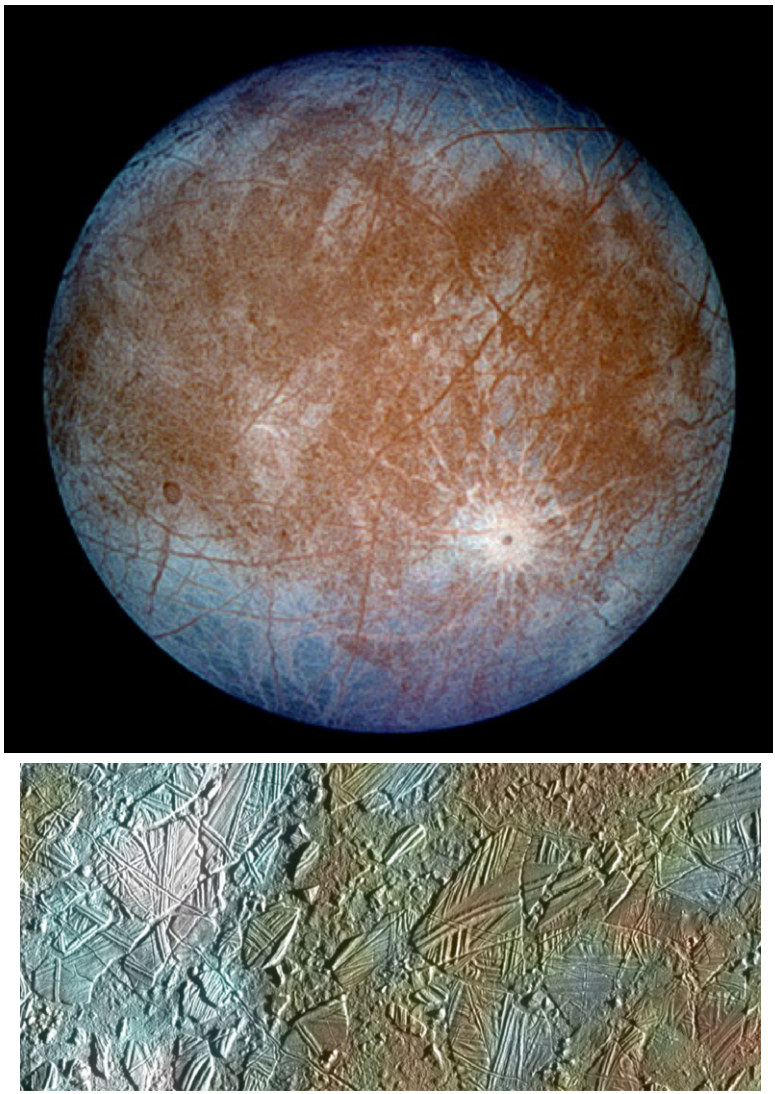


Figure 9.6 *Top*: The trailing hemisphere of Europa as imaged by the Galileo spacecraft of NASA at a distance of about 677,000 km in false-color to enhance differences in the predominantly water-ice crust of Jupiter's moon. Dark brown areas represent rocky material derived from the interior, implanted by impact, or from a combination of interior and exterior sources. Long, dark lines are fractures in the crust, some of which are more than 3,000 km long. The bright feature in the lower third of the image is a young impact crater some 50 km in diameter. *Bottom*: View of a small region of the ice crust of Europa. The white and blue colors outline areas that have been blanketed by a fine dust of ice particles ejected at the time of formation of the impact crater, some 1,000 km to the south. Europa is 3,122 km in diameter, a little smaller than the Moon (3,476 km). (*Credit*: NASA-JPL.)

light-years, is orbited by three giant planets and simulations indicate that, among them, probably resides a small rocky Earth-size world located in the habitable zone and capable of attracting enough water to harbor and support some form of life. A group of European astronomers using ground-based instrumentation (HARPS spectrograph on the ESO 3.6-meter telescope in Chile) have discovered an Earth-like planet 20.5 light-years away orbiting a small red star called Gliese 581 (see Box 9.1). The planet lies in the middle of the habitable zone, with average surface temperatures estimated to be between zero and 40°C. It could be covered with rivers, lakes and even oceans. It may be the best candidate so far for supporting extraterrestrial life in our neighborhood [26].

Box 9.1 Gliese 581-c

The red star Gliese 581, already known to harbor a 'hot Neptune' (581-b), has been found to also possess at least two extra solar super-Earth planets. One, (581-d), has a 8.2 Earth mass and orbits at 0.25 AU from the star. Gliese 581-c has a mass of 5.1 Earth masses and resides in the habitable zone of the star. It resembles our own Earth: it has the right temperature to allow liquid water on its surface, and its diameter is about one-and-a-half times that of the Earth. The range of temperatures has been estimated to be very similar to those on Earth. It probably has a substantial atmosphere similar to ours and may be covered with large amounts of water.

It is not clear what this planet is made of. If it is rock, like the Earth, then its surface may be land, or a combination of land and ocean. The surface gravity is probably around twice that of the Earth. It orbits Gliese 581 at a distance of only 0.073 AU (11 million km) and its 'year' only lasts 13 of our days [26].

Some of these extra-solar Earth-like planets might also host advanced civilizations having reached a high level of technical maturity, even though this is difficult, if not impossible, to accurately forecast today. The number of these, however, seems to be small but not too small: a recent estimate gives one civilization per volume of a few thousands light-years in our galaxy, that is a maximum of a few hundred in total [27]. Sometime in this century it might well be possible to develop the proper kind of high-resolution imaging devices that will allow us to observe surface details on the closest, such as oceans and continents [28]. We may easily guess that very few if any of these planets will be in a state similar to ours, but the discovery of oceans and continents will offer new perspectives in the search for extraterrestrial life. To consider that we might be able to visit one of these objects and to settle on any one in the future is another problem, as they lie several light-years away, and any attempt of reaching and possibly terraforming them should first solve the issues of interstellar travel.

9.2.4 Interstellar travel

Traveling through the Solar System and outside through the interstellar medium has been one of the most popular dreams of science fiction books for a long time. It is indeed a genuine dream, with little connection to the real conditions that would need to be faced when considering leaving our Solar System. The problems are fairly easy to understand. They are related to the distance of the star and to our present propulsion technologies. The nearest star to the Sun is Alpha Centauri, 4.3 light-years away from us. On 15 August 2006, nearly 30 years after its launch in 1977, NASA's probe Voyager 1, one of the fastest traveling spacecraft ever launched, with a velocity of 17 km/s, nearly 0.006% the speed of light, had reached a distance from Earth of 100 AU, approximately 14 light-hours, and is now progressively leaving our Solar System. It would take about 70,000 years at that velocity to attain our nearest stellar neighbor. Current chemical propulsion technologies are therefore not fast enough to reach the nearby stars in a reasonable time, unless new techniques are developed.

In that domain, science fiction literature is full of imaginative concepts, using nuclear and fusion, matter–antimatter engines, photon-pushed sail systems, etc., the ultimate goal being to reach a substantial portion (a few percent) of the speed of light [24, 29]. With its higher energy/mass ratio, nuclear power offers the best prospect for achieving this goal, and concepts of nuclear engines have been, and are, under study – in particular, in US industry under NASA contracts. However, the prospects are for reaching, at best, only a few thousandths of the velocity of light, placing Gliese 581-c at some 4,000 to 20,000 years away, much longer than a human lifetime. We are still far from the ultimate goal of landing humans on an extra-solar planet!

One key danger for the travelers on board these spaceships as soon as they leave Earth is space radiation, in particular the bombardment by high-energy Solar Proton Events and cosmic rays which may trigger cancers, cataracts, bone loss hereditary effects and neurological disorders [30]. On Earth, the thickness of the atmosphere provides an efficient shield against these lethal particles. On Mars and on the Moon, adequate solutions must be found, probably building underground facilities. In interplanetary or interstellar space, the human body would be submitted to a flux of about 5,000 ions every second [30, 31]. NASA-led studies estimate that about one-third of a human body's DNA would be destroyed each year spent in deep space. These estimates are subject to major uncertainties due to the poor knowledge of the biological effects of charged particles and to the characterization of the space radiation field. In the case of a trip to Mars it is estimated that one astronaut out of 10 would die of cancer after one year of exposure, not taking into account all kinds of other serious diseases such as damage to the brain cells. This requires installing absorbing protective shields in the spaceship. The best option would be to use the onboard water tanks, since water will be a necessary component of the vessel. The minimal estimated thickness of these walls would be about 5 meters which, unfortunately, should raise the minimum weight of the spacecraft to some 500 tons [30]. For the sake of comparison, the present payload capability of the US space shuttle

is only 30 tons. Other more elegant – but probably more risky – solutions, such as using magnetic shields acting like a mini-magnetosphere around the vessel, are under study in the USA and in Europe [32]. This clearly puts the problem in perspective and illustrates the need for more powerful space transportation systems in the future if it is ever envisaged to embark on such ambitious journeys.

Hence, the (naive) dream of Tsiolkovsky seems to remain just a dream at this stage. The farthest distance from Earth man has ever been is one light-second away (the Moon). The astronauts on the International Space Station are circling the Earth at a little more than a thousandth of a light-second. The nearest stars are still the same distance away as they were in 1957 when Sputnik 1 was launched since we have not invented a faster rocket than the Semiorka that sent Sputnik 1 into orbit.

Nevertheless, we are free to imagine that one day, within the next centuries or the next 100,000 years, we may be able to build a spacecraft that can travel at a velocity close to the velocity of light. Indeed, relativistic time dilatation would make the voyage seem much shorter for the travelers. Unfortunately, they would soon realize that more time had elapsed on Earth than they feel had passed. Houston might no longer respond to a 'problem' on board. Those who had conceived the mission and who had been in charge of its control, would be long since dead and no longer be in a position to respond or claim 'mission success'. For their occupants, these spaceships will have to be gradually and completely autonomous, if not just for the fact that communications between them and the Earth will make it impossible to react to problems in real time and, the further the spacecraft, the more demand for energy on board to transmit any signal back to Earth. Without mentioning the problems posed by the obsolescence of technology which makes it difficult, even today, to read the old magnetic tapes of the 1960s with present CD readers. Communications may be impossible between the Earth and the spaceship launched several centuries in the past!

Autonomy is certainly coherent with the concept of leaving Earth definitely. The ship, or rather the ships, should have enough resources in power, water, food, etc., to sustain a substantial portion of the population of the Earth. It should contain enough genetic material to maintain diversity through generations, not only for humans but also for animals and plants – in other words, for maintaining some kind of biodiversity. These vehicles will represent the best approximation to Noah's Ark in the age of space emigration. Their occupants should organize their proper education system to ensure that their descendants will acquire the know-how to repair or improve their machines, maintaining a minimum level of self-sufficiency. Again, we enter the realm of science fiction. Beyond the orbit of Mars, the 'dream' may indeed soon become a nightmare for those brave people who were naive enough to venture into these deadly and inhospitable territories where survival will never be easier than on Earth, where resources will be finite, where distances will ever be increasing, with the only perspective offered of living in a limited vessel with no biodiversity, no culture, no family, and probably no food!

9.2.5 Space cities?

At the time when Tsiolkovsky thought of leaving the cradle of the Earth, a little more than 100 years ago, Einstein had not yet invented his theory, the Universe was static and its age evaluated at a few million years only, and the population of the Earth was less than 2 billion people. Hope was unlimited, everything was possible! Some 80 years later, an interesting concept was proposed in 1969, in the wake of the Apollo missions to the Moon, by the leading proponent of space colonization G.K. O'Neill, also professor of physics at Princeton University [33]. He has been a source of inspiration for a lot of fiction books and movies, which is probably where his ideas have found most use. They also might have offered an opportunity for students in engineering to exert their imagination and skills. However, they lack a proper system analysis, involving all aspects of such complex and gigantic space stations, an essential element for designing and developing any space mission. Such utopian proposals often originate in the mind of engineers – and sometime of scientists – who have never built or launched a satellite. Their seriousness can be assessed *a posteriori*, nearly 40 years after Apollo, looking at the intended original goals and at where we stand today.

The original claim in the confidence of O'Neill's concept was that: '*Careful engineering and cost analysis shows we can build pleasant, self-sufficient dwelling places in space within the next two decades (i.e. before 1994)*', solving many of Earth's problems. A first space colony of 10,000 people would have been in place in 1988 (i.e. two years after the Challenger accident!) and colonies between 200,000 and 20 million people would be able to live in such habitats in 2008, the year when this book is published! The model would lead to establishing by 2050 a space population of about 14 billion and decrease that remaining on Earth from a maximum of 16 billion (according to his estimate) to a stable (?) level of 2 billion onwards! In comparison with the real world, we are painfully assembling the International Space Station which is, on average, occupied by no more than a handful of astronauts, and the operational costs of the station are so high that they may lead to its abandonment.

A very modest example of something akin to O'Neill's idea, but on Earth, was the Biosphere project in Arizona – an experiment for a sustainable, isolated human outpost. Surely doing this on Earth would be simpler than in space. Unfortunately, the project has been abandoned because it was judged much too expensive and required enormous amounts of power for maintaining adequate living conditions underneath the dome. Survival was indeed the main activity of the crew of 8 who volunteered to participate in the project, as the life-support system is constantly put into question. The fact that the crew emerged from their two-year closure still speaking to each other and apparently in better health than when they started, was considered at that time to be an accomplishment [34]!

Our little intrusion in this universe of dreams shows that space emigration in both ways, from the Earth to other solar systems and vice versa, is probably impossible. This may provide an answer to Fermi's paradox. In 1950 in Los Alamos, at a lunch discussion about flying saucers and extraterrestrial life involving a group of atomic scientists – among whom were Edward Teller, the

Hungarian physicist 'father' of the American hydrogen bomb, Herbert York, an American nuclear physicist, and the 1938 Physics Nobel laureate Enrico Fermi himself – there was a consensus that the Universe should contain billions of planets capable of supporting life, and most probably millions of intelligent species. Fermi made a rapid calculation and remarked that these putative civilizations, based on the human tendency to expand and on the promising capabilities of space technology, would already have colonized the entire galaxy within a few million years and visited us a long time ago and many times over. He then asked the stunning question: *'But where is everybody?'* This is since then known as the Fermi Paradox [35]. Even though the duration of any intelligent civilization is very small with respect to that of its parent star, reducing the probability of having several civilizations contemporaneously habiting our galaxy, we dare to suggest that the simplest answer that we might offer to Fermi's question is that 'everybody is at home' living in splendid isolation because, unfortunately, as we have discussed, interstellar travel is just an impossible concept.

This conclusion should not be interpreted as meaning that we will not expand and explore other words in the future, in particular the Moon and Mars, which are within a reasonable distance. If we do this, however, it is our feeling that it will not be as an entire civilization, but rather as explorers, with the goal of maintaining or improving our living conditions on Earth, through acquisition of scientific knowledge, or serving essential equipment in orbit or on the Moon, as was done so successfully with the Hubble Space Telescope, or searching new or exhausted Earth's resources if proven economically viable, or just for fun – *'pour le sport'* said former French minister of research and Academician, Hubert Curien [36] – or just as sheep in the flock, if one goes there the others will follow.

9.3 What to do with the Moon?

The Moon is probably the most symbolic target of human space exploration and the astronauts of the Apollo program as well as the Soviet Luna robots have clearly demonstrated that we can go there, land there, walk there and return. The Moon, at only one 'light-second' distance, no more than three days away from Earth by the present means of transportation, does not present the same challenges as the long journeys to Mars, or further out in the Solar System, that we just discussed. The prospects of technical progress in avionics and rocketry open many possibilities to return to the Moon more often and at lower costs than before. It is presently the selected target of all space organizations or political entities that want to impress their governments or constituencies, and start a new era of space exploration. The two Bush Presidents of the United States have made the Moon the focus of their exploration initiatives. All major space agencies have at least one Moon mission in their program. Most are unmanned, such as the Japanese KAGUYA (Figure 9.7), but plans also exist to send again American and, in the near future, Indian and Chinese astronauts. More and

more, the Moon seems to be considered as a suburb of the Earth, part of the Earth system. But what will we do on the Moon or what will we do with it? Several studies have been undertaken to provide an answer to these questions mostly in Europe by ESA [37], and in the United States by NASA and the American Academy of Sciences [38].

9.3.1 The Lunar Space Station

With a diameter of 3,476 km the Moon is a relatively small body. Because of its low gravity, only one-sixth that of the Earth, it is delicate to conduct large-scale activities on it. The landings of spacecraft and the presence of manned stations will have a long-lasting impact on the environment, in particular on the natural 'atmosphere' of the Moon as well as on its surface [37]. It has been estimated that the successive Apollo missions, plus the other American and Soviet landers of the 1970s, have left some 100 tons of exhaust gases on the Moon. It took 24 years after the last astronauts left to return the environment to the pre-Apollo state [39]. Future permanent activities may well release gases in amounts that will surpass the natural influx coming from the solar wind and the meteoritic bombardment.

The dust problem is certainly one of the most serious challenges to all those who envisage working on the Moon. Both robotic and human activities would generate clouds of dust particles from the thin regolith that may remain in orbit and would gradually fall back and spread all over the surface, and on equipments in operation. The dust is electrostatically charged and will stick to all unprotected parts, necessitating regular cleansing and maintenance. From a 'pure science' perspective, the detailed local chemical composition of the pristine regolith will definitely be blurred by the dust falling back and will lose part of its original message. Of course, this argument may not hold in case it is definitely decided that the scientific assets offered by the Moon may be sacrificed for the benefit of industrial and/or touristic exploitation. But even in that case, dust pollution remains a real problem as it is also feared to be toxic to humans. Furthermore, the surface may be radioactive as a result of the high-energy particles from the solar wind impacting the rocks and the regolith.

Besides the dust, the Moon presents other challenges. Because of the 14-day night, solar energy cannot be used continuously. Storage devices do not necessarily offer the best solution and recourse to nuclear power is probably necessary, as envisaged by NASA in their Constellation program. As shown on Table 9.2 [44], accessing the Moon's surface requires additional propulsion energy, as 11.2 km/s are required to go from Earth to free space plus 6.5 km/s to descend to the lunar surface. The Moon can host small colonies of scientists or engineers, as is done in the stations on Antarctica or the off-shore platforms, depending on what it is envisaged can be done on the Moon: research, exploitation of resources or, as quoted more and more often, tourism! Any one of these possible utilizations is presently envisaged, in particular by the United States, even though they are probably not mutually compatible and certainly do not reflect a fully coherent approach, which is a point we now address.



Figure 9.7 A spectacular view of the Earth rising above a dark landscape of lunar craters seen in grazing solar light by the high definition imaging camera on board KAGUYA (SELENE), Japan's first large lunar mission and the largest since the Apollo program. From its 100-km altitude polar orbit above the lunar surface, KAGUYA is gathering information on the elemental and mineral composition and on the remnant magnetic field and gravity field of the Moon. These data will be used in view of further human exploration of our natural satellite. (*Credit*: JAXA/NHK.)

Table 9.2 Mission Velocity requirements for reaching the Moon and the Near Earth Asteroids. Target accessibility depends on the velocity change delta-v to inject into transfer orbit, plus the velocity change needed to rendezvous with the target. (Adapted from Sonter, reference [44].

Earth surface to Low Earth Orbit	8.0 km/s
Earth surface to escape velocity	11.2 km/s
Earth surface to Geostationary Orbit	11.8 km/s
Escape velocity from Low Earth Orbit	3.2 km/s
Low Earth Orbit to Geostationary Orbit	3.5 km/s
Low Earth Orbit to Moon landing	6.5 km/s
Geostationary Orbit to Moon landing	2.8 km/s
Low Earth Orbit to Near Earth Asteroid	~5.5 km/s
NEA to Earth transfer orbit	~1.0 km/s

9.3.2 The Moon as a scientific base

As far as science is concerned, the most interesting assets offered by the Moon are certainly the study of the Moon itself and that of the history and evolution of the Sun, of the Earth and the Solar System as discussed earlier in Chapter 2 (Section 2.3). The hidden side of the Moon also offers a unique place to do radio-astronomy in a clean electromagnetic environment since the Moon acts as a

shield against electromagnetic pollution from the Earth, which is becoming a severe problem for observations. Craters could be utilized as natural fixtures for large telescopes just as, on Earth, the Arecibo radiotelescope in Puerto Rico, installed in the crater of an inactive volcano, is used for observations of objects passing through the field of view. This is known as transit-mode observation. Because there is no atmosphere around the Moon, it might be envisaged that Arecibo-type telescopes – or other types – could be built there for wavelengths in the radio and far infrared spectral range. All other areas of astronomy are covered more easily and probably more cheaply from free space on dedicated orbits rather than on the Moon, which offers no particular specific assets. It adds to the well-known difficulties inherent to the use of space equipment the inconveniences that affect astronomy on Earth, apart from the absence of atmosphere: gravity (although much reduced on the Moon), large temperature excursions and a horizon restricting observations to one half of the celestial sphere. Furthermore, the impacts from micrometeorites and the dust will require frequent cleaning and maintenance activities involving very specialized technicians as was done for Hubble. This would not make such facilities competitive costwise when compared to free flying fully automated orbiters [37]. In the area of radiation biology, the Moon certainly offers interesting prospects for investigating the biological importance of the various components of cosmic and solar radiation. In preparation for a future human lunar outpost, radiation monitoring, shielding and solar wind shelters could be tested directly on the Moon [40].

9.3.3 The Moon for non-scientific exploitation

Lunar resources

We know from the analysis of the lunar rocks and from lunar orbiters that the soils of the Moon comprise about 20% silicon and 30% metals such as aluminum, iron, titanium and magnesium (see Table 2.2). ESA's Smart-1 orbiter has also discovered large quantities of calcium in the regolith. There is ample oxygen, representing more than 40% of the weight of the lunar soil, but it is all bound up in compounds such as silicates that are difficult to break down. Hydrogen is also present, but much less abundant, at the level of 0.001%. There is very little water with the exception perhaps of ice, which is possibly hidden in the permanently dark sides of large craters near the poles. Hydrogen could hence be used to produce water in combination with the oxygen contained in the silicates, if it is proven feasible to extract it from the lunar rocks. There is outspoken interest to use these lunar resources either on the Moon for the construction of lunar bases, or for building spacecraft to go from the Moon to outer space (Mars, for example) or to import them to the Earth. Indeed, lofting from Earth the large quantities of materials that would be required to build a lunar infrastructure, and to support any of the activities that might be considered on the Moon, be they of a scientific or industrial/commercial nature, does not

look like a very efficient process, even in the distant future. Lunar resources might, if possible, be exploited 'on the spot' for that purpose. Conversely, the lower gravity of the Moon has been considered as an advantage to export some of these resources to the Earth [41], but the problem is to come back and land on Earth (Table 9.2). This might be accomplished through an intermediate space station in low Earth orbit, at the expense of course of building and maintaining such an expensive infrastructure.

Helium 3

One of the most appreciated potential resources of the Moon is the light isotope of helium, ^{3}He. The Moon has been bombarded for billions of years by the solar wind, which consists largely of ionized hydrogen and ionized helium. ^{3}He is being formed in the very energetic solar flares and is also transported by the solar wind to the lunar surface where it remains trapped in the regolith. ^{3}He is a rare element on Earth as opposed to the heavier ^{4}He isotope, but it is found in relatively large quantities in the lunar regolith with concentrations between 10 and 20 parts per billion, but the regolith has to be heated to about 600°C to extract the precious isotope. ^{3}He has been considered as a very interesting element for solving the energy problem on Earth, exploiting its fusion with deuterium (see Chapter 7, for a discussion of nuclear fusion using the D + ^{3}He reaction). On Earth, the total available resources of ^{3}He amount to only a few hundred kilograms, mainly produced by the decay of tritium (T) in nuclear weaponry. The ^{3}He + D reaction requires higher temperatures than the T + D reaction used in ITER and therefore is more difficult to realize in a fusion reactor. When it is proved that the D + ^{3}He reactors can be developed, then the lunar resources might begin to be seriously considered. Table 9.3 provides some data on the lunar resources evaluated after the Apollo missions. Harrison Schmitt, one of the last Apollo astronauts – and a geophysicist – estimates that about 2 km^2 of the Moon's surface, excavated to a depth of 3 meters, will provide 100 kg of ^{3}He, enough to power a 1,000-MW power plant. One metric ton of ^{3}He would supply a city of 10 millions inhabitants with electric power for a year [42]. Assuming that the whole lunar surface has the same average content of ^{3}He everywhere, a rough calculation shows that a world population of 10 billion inhabitants could be fed by ^{3}He fusion-generated electricity for approximately 500 years. This is not very long with regard to 1,000 centuries but the situation is even worse if we consider the whole 63-TW power demand of the 100,000-year world, which, as mentioned in Chapter 7, would exhaust the whole ^{3}He content of the Moon in just 15 years! This is a maximum because ^{3}He is not regularly spread over the lunar surface.

Unfortunately, after that time, the whole area of the Moon would have been excavated to a depth of 3 meters. Every year, some 20,000 km^2 of the lunar surface would be excavated, which represents a mammoth public infrastructure development project that has no equivalent on Earth. According to Schmitt, digging 2 km^2 of the Moon to a depth of 3 meters requires an hourly mining of an area about 28 m^2 and the processing of the finest 50% of the mined soil or

Table 9.3 Useful data on lunar Helium-3 as found in [37] and [42]

Parameter	Value
Ratio $^3He/^4He$	0.5×10^{-3}
Concentration of helium in lunar soil	~36 g/ton
Concentration of ^3He in lunar soil	~13 mg/ton
Depth down to which ^3He is distributed	~3 meters
D^3+He energy equivalent of these resources	2×10^6 GW years

2,000 tons, to extract its volatiles. That would be able to feed only 1 million people! On top of this, the precious resource must be transported down to Earth after being processed on the Moon, demanding a fairly sophisticated industrial infrastructure. In the mind of Schmitt, mining lunar ^3He would not eliminate other sources of energy production. Nevertheless, Moon mining, at any fraction of what would be necessary to feed the Earth's population (if it one day occurs), would be a very expensive project. It would in addition irremediably destroy a substantial portion, if not all, of the lunar surface while not being able to provide a solution to the energy problem for more than 15 years in an optimistic evaluation. Trading 4.5 billion years of Solar System history for 15 years of energy maximum (500 years for present-day electric power) is not a good deal, since we need much more energy for many more years! As for the Earth, lunar resources are limited and not enough to sustain the needs of 11 billion people for 1,000 centuries. This does not look like an adequate solution! Furthermore, we do not need it. When fusion becomes operational, it will be much easier to produce energy using lithium, which is very abundant on Earth in the D + T reaction (Chapter 7), than by excavating the first meters of the Moon's surface.

Exotic ideas

Other exotic but nonetheless potentially useful ideas have been imagined for making good use of our natural satellite [37]. One of them would consist of creating a long-duration archive on the Moon, as many records of the development of humanity and civilizations have been destroyed for ever due to environmental catastrophes or deliberate human actions and wars. The longevity of the Moon is about 4 billion years, much longer than our 100,000-year period and allows the storage and archiving of all these records for as long as the Moon is not swallowed by the Sun.

Another more immediate possibility would be to use the Moon as a recycling facility for materials used in the manufacture of satellites residing either in Low Earth Orbit (LEO) or in the already crowded geostationary orbit, thereby liberating that orbit – which is so vital for our future – from useless dead spacecraft. It is the most crowded orbit of all used by artificial satellites, be it for telecommunications, meteorology or Earth observation in general. Since the 50 years following the launch of Sputnik 1, more than 4,500 satellites have been

launched, of which approximately 40% are using that orbit. As these satellites are essential for the management of the Earth (see Chapter 10), it is expected that many more will be launched in the future. At the present rate of about 20 launches per year – and considering that this number will increase as more countries will need them – it is easy to guess that a further 2,000 satellites will occupy the geostationary orbit per century, or more than 2 million over 100,000 years. After their mission is finished they are there for ever unless the organizations to which they belong obey the present guidelines of liberating the orbit (see Chapter 3). Furthermore, they contain precious materials that would deserve to be recycled. As indicated in Table 9.2, it needs just an extra velocity of 2.8 km/s to reach the Moon's surface from geostationary orbit, which is less than would be required to reach that orbit from LEO. This shows that an easily accessible recycling facility on the Moon could be envisaged in the future which would allow the re-use of precious resources that might be lost for ever in space. Of course, the devil is in the details of such a concept and is certainly not easy to implement. But is it more difficult than extracting the same resources from the Moon? The surface of our satellite might also be used to install some of these essential Earth observation satellites that do not absolutely require the use of the geostationary orbit for the accomplishment of their mission.

The conclusion at this point is that we will most likely go back to the Moon, be it for scientific research or for tourism, for extracting resources or to access facilities that are seen to be essential for the survival of humanity. This blend of interests would justify an important effort of international cooperation on a broad basis, certainly involving the main space nations of the world. A worldwide initiative aimed at creating a consensus among the scientists, the politicians and the space organizations is therefore a necessity. Since these activities may not necessarily be mutually compatible, there is a clear need for a legal basis to regulate them, as in the case of Antarctica, through the Treaty adopted in 1959. Since then, the expansion of activities there has been entirely justified on the basis of science and in the framework of international cooperation. The Space Treaty of 1967 is rather flexible about the issues that concern the rights of lunar or Martian explorers. It is restrictive only on the control of property rights, just as the Law of the Sea prevents anyone owning the sea. The appropriation of resources is not forbidden by the treaty but some rules of the road must be established that go beyond the terms of the treaty in order to prevent deteriorating a unique scientific asset and avoiding what might one day resemble a 'lunar war'.

9.3.4 Resources from outside the Earth–Moon system: planets and asteroids

Other bodies of the Solar System possess resources that are either in greater quantity than on the Moon or are easier to extract. If ^{3}He is still considered as an essential alternative to the fusion reaction of D + T, the atmosphere of Uranus offers yet another for extracting the precious gas [43]. However, the practical mounting of such an ambitious concept would involve an impressive number of complex subsystems and operation, such as, among others:

- A space station in orbit around the Earth.
- 'Slingshots' through Jupiter and Saturn gravity wells.
- A high-speed Uranus atmospheric entry probe.
- Parachutes to slow the probe's velocity.
- Cylindrical balloons to 'mine' the atmosphere of the giant planet.
- A megawatt nuclear reactor to heat the air and flush the balloons with heated air.
- A liquefier of the atmosphere to extract helium-3.

In addition, the complexity of returning the cargo to Earth is phenomenal as the number of flights needed to feed the whole world with energy has been evaluated to be equivalent to 300-ton loads per month to be transported from Uranus to the Earth via an Earth re-entry system! Even if we anticipate a drastic reduction of the space travel costs in the centuries to come, it seems highly improbable that this solution is economically viable. Furthermore, we have just seen that in fact we do not need ^{3}He.

Asteroids, on the other hand, offer a more realistic alternative for resources, in particular the Near-Earth Asteroids (NEAs) that have orbits close to that of the Earth (Chapter 3) [44]. It is evident from the numbers in Table 9.2 that the NEAs are not too difficult to access: approximately 10% can be reached with much less extra velocity than it takes to reach the Moon. This is true also for the return velocity back to Earth. Landing on a NEA is also not too difficult, as we have seen in Chapter 3.

Asteroid geology allows us to establish reasonable correlations between the different spectral and photometric properties of these objects and their inferred surface composition, although only a small number have yet been spectrally classified. They represent potential sources of material, either for the development of interplanetary infrastructures (perhaps!) or as replenishment sources for any of the resources that would be missing on Earth in the future. At least 50% of NEAs are likely to be promising ore bodies. For example, the smallest Earth-crossing asteroid, 3554 Amun, is a 2-km piece of iron, nickel, cobalt, platinum and other metals; it contains 30 times as much metal as we have mined throughout history on Earth, although it is only the smallest of dozens of other known metallic asteroids.

Hence, mining asteroids can be envisaged in the future if it became absolutely necessary and if we had no other possibilities for extracting the precious resources directly from the Earth, as whatever we plan to mine in space will certainly be more expensive than from the ground of our planet. Such activities do not necessarily need to be manned, however, as tele-operation and clever robotics can ensure the success of such enterprises. Prior to that, it is advisable to accurately evaluate the detailed chemical composition of the most easily accessible objects, either by remote-sensing techniques or *in situ*, through landing missions. It would therefore be important to have a better understanding of landing, anchoring and de-anchoring techniques, as well as of soil processing through small automated mining test stations for several typical targets, as all may not present the same properties.

9.4 Terraforming the Earth

At this point, it appears that leaving Earth to secure the future of humanity stays in the realm of dreams or utopia. Terraforming Venus, Mars and *a fortiori* Titan and Europa have been disregarded because the transformations are drastic, with no guaranteed success, even during the time frame we consider here. Besides the gigantism of such projects, their costs, timescales and decision-making processes dictate that we cannot seriously consider them as solutions. Therefore, the Earth remains the best approximation that we have today of a habitable planet, even though we have already started to modify its environment, its climate, and are exhausting some of its essential resources. Placing our future on Earth seems at this stage our only credible and viable option, but this requires the proper management of the planet – an issue that we address in the last two chapters of the book.

The concern expressed in Chapter 6 as to the evolution of the climate is now broadly shared among a large number of scientists, politicians and the world population. Rather than sticking to the idea of re-engineering Venus and Mars, the concepts of Earth re-engineering, what we call 'Earth Terraforming', are being proposed by some famous and serious people and may be worth considering. The issues mostly concern global warming and how to stop its trend while continuing to burn fossil energy by, essentially, lowering the greenhouse effect or cooling the Earth. We briefly review those concepts now.

9.4.1 Absorbing or storing CO_2

The first category of concepts attempts to limit or decrease the quantity of greenhouse gases in the atmosphere, in particular CO_2. In 1976, Freeman Dyson [45], the same person who evaluated how to re-hydrate Venus by importing hydrogen from Uranus (see page 286) proposed this time to purge the Earth's atmosphere biologically, using intensive rapid-growth plantations of either trees or swamp-plants such as water hyacinths, and converting them into humus or peat. The side-effect of this remedy would be a very substantial demand for fertilizers that would introduce elements in the soil such as phosphorus, nitrogen and potassium, creating a new imbalance and limiting the possible scale and speed of the project. Concerning methane, the other greenhouse gas, the use of bacteria or archea would probably leave the Earth in a state much worse than that we would like to correct.

Another concept, not necessarily less risky, would be to seed the oceans with iron particles to allow phytoplankton to absorb the extra amount of CO_2. Such experiments have been attempted in the Pacific Ocean, and satellite imagery shows indeed an increase of the phytoplankton after iron seeding, but nothing proves that in the end the carbon has been deposited at the bottom of the ocean and has not been returned rapidly to the atmosphere. Furthermore, collateral effects, such as the formation of anoxic areas in the ocean and the proliferation of bacteria capable of degrading nitrates, subsequently releasing into the atmosphere nitrogen protoxide (N_2O) – another greenhouse gas – have not been

estimated because it is very difficult to do so. They may well be quite detrimental to the environment.

More seriously, following the commitments made by the Kyoto Protocol, the concept of capturing CO_2 and store it underground is implemented on an experimental basis in Europe and in the United States. That approach would offer the important advantage that would allow the continued use of fossil energy without worsening the climate system further. The basic idea is that CO_2 is captured during the production process and stored safely underground or in the deep ocean so that it can dissolve in sea water for a long period. Empty oil and gas reservoirs, coal seams and porous rock beds can be used for such storage and according to the IPCC, the available capacity would be enough to hold around 40% of the world emissions. The European Union have selected the CO_2 SINK project which consists of injecting some 60,000 tons of CO_2, which corresponds to the annual output of some 40,000 cars, into a saline aquifer 700 meters underground near the town of Ketzin, west of Berlin [46]. Another project jointly envisaged by India and the United States is to use basalt deposits in the northwest Pacific states where 50 Gt of CO_2 could be stored or in the Deccan Traps which might be able to hold 150 Gt [47]. All these projects are not without risks, and worries are that the stored CO_2 would seep out and not stay where it was originally stored, leaking carbon back for decades or centuries. Under the sea, the reservoirs might also eventually trigger landslides and tsunamis. Therefore, the choice of the storage sites appears to be very crucial. In spite of the intrinsic difficulties and problems of the concept, it certainly remains a credible option, if not the only one, to eliminate atmospheric CO_2 while continuing to burn fossil fuels, hence, it is essential to assess it more in depth through, for example, pilot projects on a small scale.

9.4.2 Cooling down the Earth

Some serious scientists are also proposing to stop global warming by cooling down the Earth's temperature. One of them is Nobel Prize winner Paul Crutzen who proposes to produce artificial aerosols made of small amounts of sulfate particles (1–2% of the amount that is emitted in the troposphere), and release them in the stratosphere. They would reflect some of the incoming solar radiation and cool the planet by some 1 to 2°C [48]. The 1991 Pinatubo eruption offered a 'natural' test of this effect when it injected into the atmosphere large quantities of sulfur dioxide that rapidly created a veil of aerosols, lowering the global ground temperatures by 0.5°C on average over two years (see Chapter 4).

This idea presents of course potential risks that have to be thoroughly analysed, and much research is needed to ensure that no major environmental side-effects would exist [49]. Aerosols injection might perturb the whole atmospheric–oceanic circulation, in particular the natural Arctic oscillation phenomenon, inducing local warming effects in wintertime and the reverse in other places. Indeed, the winter following the Pinatubo eruption witnessed a very marked cooling in nearly all the parts of the world except northern Europe,

where a warming occurred. Furthermore, cooling by aerosols is only efficient when the Sun is shining, while greenhouse gases are efficient day and night, with a higher effect at high latitudes. The process also leads to inhomogeneous effects not easy to assess region per region [50]. Furthermore, the stratosphere is not a static layer: it is a complex system in itself implying physics, chemistry and a lot of circulation and interaction with the troposphere. With such projects, the whole climate system is at stake. Their complexity, therefore, requires a thorough scientific analysis involving, for years probably, many specialists from climatology, oceanography, geology, astronomy, biology, agronomy, etc., a work in essence very similar to that of the IPCC. In fact, the possibility that the IPCC should create a special task force to look at all the positive or negative effects is one to be considered with serious attention.

Another geo-engineering concept also discussed earlier in the case of terraforming Venus, is to block the sunlight far above, from space, by shielding the Earth, subjecting it to the permanent shadow of a screen or a swarm of small satellites. A sun shield a little larger than the Earth at Lagrange point L_1 would cut down the solar constant on Earth by about 1.8% [51]. Unfortunately, as in the case of Mars, the project seems to be totally gargantuan and unrealistic for many reasons: the number of fliers required has been evaluated to be about 16 trillion and the cost in US dollars one-third this number. Not to mention the operational aspects of maintaining the shade at L_1 to secure a constant amount of dimming!

We may question at that point why several serious people in the scientific world are exposing themselves so openly – and so imprudently – inventing such unrealistic and potentially dangerous projects. Some are honestly concerned about the present situation and are keen to avoid the risk of last-minute bad surprises. They apparently use that means as a signal or an alert for their colleagues and the political world. Others certainly find intellectual satisfaction in defining large-scale engineering projects but do not care too much about their realism. And some others are still navigating between science and fiction! Our feeling is that none of these projects is at this time realistic, and none has been seriously studied. They give a poor image of the Earth, making it look like a wounded person on whom the physicians and the surgeons realize that something has to be done to save it, but do not know exactly what to do and are ready to experiment with anything. The Earth is such a complex system and its climate is a manifestation of that complexity; therefore, it is neither wise nor prudent to add new elements or parameters to that complexity such as the introduction of artificial aerosols or a permanent sunlight dimming – not to mention the bioengineering concepts discussed above. The best alternative at the present time is certainly to take the proper actions to stop the trends in CO_2 emissions as discussed in the previous chapter. In case there is no other choice, the decision process to undertake such projects must be political and global, assuming a worldwide consensus and the assurance of the necessary continuity. This is particularly true in the case of the aerosol-cooling project – the only one that would be worth studying in more depth. It is obvious that the whole

population of the world is concerned and the decision to implement it is one for either the United Nations or the new governance we discuss later in Chapter 11. None of these conditions is assembled at the time of printing this book, but in an optimistic scenario they might be in the future, and hopefully not too late!

9.5 Conclusion

This discussion leads us to the inevitable conclusion that we are bound to the Earth for the next 100,000 years. There is no serious alternative to our occupying the mother planet for another 1,000 centuries, even if we manage to inhabit the Moon, Mars and perhaps Titan, but it would just be for the same reasons that we inhabit Antarctica today: for science, resource exploitation or tourism! It may be disappointing that the dream of Konstantin Tsiolkovsky cannot be seriously considered as the ultimate solution to securing our future. Contrary to what was thought and imagined by all the advocates of space colonization, satellites and space cities are not offering humanity the remedy to all our problems on Earth. The Earth is what we have and we must make the best use of it without furthering its deterioration. That being said, space offers a different perspective. It is one of the most precious tools we have to secure our future possibly for 1,000 centuries more, and the fathers of space conquest must be acknowledged as we do now master that tool. This is what we want to illustrate now.

9.6 Notes and references

[1] According to McKay, C.P., 1982, *Extrapolation* **23,** No. 4, Kent State University Press, 309–314, the object of terraforming is to alter the environment of another planet to improve the chances of survival of an indigenous biology, or in the absence of an indigenous biology, to allow habitation by most, if not all, terrestrial life forms.

[2] Kasting, J.F. et al., 1988, 'How climate evolved on the terrestrial planets', *Scientific American* **256**, 46–54.

[3] The albedo A of a planet determines the radiative equilibrium temperature T_e that it would reach due to the absorption by its ground of direct solar illumination, and emission to outer space of infrared radiation from the planet's surface (see Chapter 2). T_e is related to the albedo and the solar constant S at the planet orbit through the formula:

$$\sigma T_e{}^4 = S/4\ (1 - A)$$

where σ is the Stefan–Boltzman constant equal to 5.67×10^{-8} W/m^2/K^4.

[4] Lunine, J.I., 1999, *Earth: Evolution of a Habitable World,* Cambridge University Press, p. 319.

[5] Bertaux, J.L., 2006, 'Solar variability and climate impact on terrestrial

planets', in *Solar Variability and Planetary Climates* (ISSI Book Series No. 23), Calisesi Y. et al. (eds), Springer Publication; and *Space Science Reviews* **125**, Issue 1–4, 435–444. Taylor, F., 2006, 'Climate variability on Venus and Titan', *Solar Variability and Planetary Climates* (ISSI Book Series No. 23), Calisesi, Y. et al. (eds), Springer Publication; and *Space Science Reviews* **125**, Issue 1–4, 445–455.

[6] Ingersoll, A.P., 2007, 'Express dispatches', *Nature* **450**, 617–618.

[7] Kasting, J.F. and Catling, D., 2003, 'Evolution of a habitable planet', *Annual Review of Astronomy and Astrophysics* **41,** 429–463.

[8] Fogg, M.J., 1987, 'The terraforming of Venus', *Journal of the British Interplanetary Society* **40**, 551–564.

[9] Dyson, F., 1989, 'Terraforming Venus', *Journal of the British Interplanetary Society* **42,** 593–596.

[10] Dyson, F., 1966, 'The search for extraterrestrial technology', *Perspective in Modern Physics*, Interscience Publishers, New York, 641–655.

[11] Sagan, C., 1994, *Pale Blue Dot,* Random House edn, New York, p. 429.

[12] Ecopoeisis has a more modest aim than terraforming. It refers to the fabrication of a self-sustaining ecosystem on a lifeless planet. The expression is derived from the Greek roots, οικος, an abode, house or dwelling place (from which we also derive 'ecology' and 'economics') and ποιησις a fabrication or production (from which we derive 'poesy', as well as a variety of other biological terms such as biopoiesis, haematopoiesis, etc.). Ecopoiesis is now used in the literature to describe the implantation of a pioneering and, hence, microbial ecosystem on a planet, either as an end in itself or as an initial stage in a more lengthy process of terraforming (see Haynes, R.H., 1993, 'How might Mars become a home for humans', *The Illustrated Encyclopedia of Mankind*).

[13] McKay, C.P. et al., 1991, 'Making Mars habitable', *Nature* **352**, 489–496.

[14] Forget, F., Costard, F. and Lognonné P., 2003, *La planète Mars, histoire d'un autre monde,* Pour la Science Eds, Berlin, p. 144.

[15] Poulet, F. et al., 2005, 'Phyllosilicates on Mars and implications for early Martian climate', *Nature* **438**, 638.

[16] Lundin, R. et al., 2005, 'Planetary magnetic fields and solar forcing – critical aspects for the evolution of the Earth-like planets', *Geology and Habitability of Terrestrial Planets* (ISSI Book Series **24)**, Fishbaugh K. et al. (eds), Springer Publication; and *Space Science Reviews* **129**, Issue 1–3, 245–278.

[17] Kurahashi-Nakamura, T. and Tajika, E., 2006, 'Atmospheric collapse and transport of carbon dioxide into the subsurface on early Mars', *Geophysical Research Letters* **33**, L18205, p. 5.

[18] Laskar, J. et al., 2002, 'Orbital forcing of the Martian polar layered deposits', *Nature* **419**, 375–377. Laskar, J. et al., 2004, 'Long term evolution and chaotic diffusion of the insolation quantities of Mars', *Icarus* **170**, 343–364.

[19] Montmessin, F., 2006, 'The orbital forcing of climate changes on Mars', *Solar Variability and Planetary Climates* (ISSI Book Series No. 23), Calisesi Y. et al. (eds), Springer Publication and *Space Science Reviews* **125**, Issue 1–4, 457–472.

[20] Fogg, M.J., 2005, 'On the possibility of terraforming Mars'. http://www.redcolony.com/
[21] Lovelock, J.E. and Allaby, M., 1984, *The Greening of Mars*, Warner Brothers Inc., New York.
[22] Oberg, J.E., 1981, *New Earths*, New American Library Inc., New York.
[23] Gott III, J.R., 2007, 'Why humans must leave Earth', *New Scientist* **2620**, 51–54.
[24] Fogg, M.J., 1991, 'Terraforming, as part of a strategy for interstellar colonization', *Journal of the British Interplanetary Society* **44**, 183–192.
[25] Raymond, S.N. et al., 2006, 'Predicting Planets in known Extrasolar Planetary Systems.III Forming Terrestrial Planets', *Astrophysical Journal* **644**, 1223–1231.
[26] Udry, S. et al., 2007, 'The HARPS search for southern extra-solar planets XI. An habitable super-Earth (5 Earth masses) in a 3-planet system', *Astronomy and Astrophysics* **469** (3), Letter L43.
[27] Cole, G.H.A., 2006, 'Observed Exoplanets and Intelligent Life', *Surveys in Geophysics* **27** (3), 365–382.
[28] Labeyrie A, 1996, 'Resolved imaging of extra-solar planets with future 10–100 km optical interferometric arrays', *Astronomy and Astrophysics Supp. Series* **118,** 517–524.
[29] Crystall, B., 2007, 'Engage the antimatter drive', *New Scientist* **2620**, 62–65.
[30] Hamilton, S.A. et al., 2006, 'A murine model for bone loss from therapeutic and space-relevant sources of radiation', *Journal of Applied Physiology* **101,** 789–793; DOI:10.1152/japplphysiol.01078.2005.
[31] Parker, E.N., 2005, 'Shielding space explorers from cosmic rays', *Space Weather* **3** (8); Parker, E.N., 2006, 'Shielding Space travelers', *Scientific American* **294** (3), 22–29; Parker, E.N., 2006, 'Peut-on protéger les Voyageurs Spatiaux?', *Pour la Science* **343**, May.
[32] Cucinotta, F.A. and Durante, M., 2006, 'Cancer risk from exposure to galactic cosmic rays: implications for space exploration by human beings', *Lancet Oncology* **7**, 431–435.
[33] O' Neill, G.K., 1974, 'The colonization of space', *Physics Today* **27** (9).
[34] Odum, P., 1996, 'Cost of living in domed cities', *Nature* **382**, p. 18.
[35] Webb, S., 2002, *Where is Everybody?,* Copernicus Books, Praxis, New York, p. 288.
[36] Hubert Curien: Private Communication.
[37] *Mission to the Moon,* 1992, ESA SP-1150, p. 190, and *Towards a World Strategy for the Exploration and Utilization of our Natural Satellite,* 1994, ESA SP-1170, p. 167.
[38] *The Scientific Context for Exploration of the Moon: Final Report,* 2007, Committee on the Scientific Context for Exploration of the Moon, National Research Council, ISBN: 978–0-309–10919–2, p. 120.
[39] Kopal, Z., 1974, *The Moon in the Post-Apollo Era,* Reidel Publ. Co., Dordrecht, Holland, p. 223.

[40] Bonnet, R.M., 1996, 'How might we approach a major lunar programme?', *Advances in Space Research* **18** (11), 7–13.
[41] The Director of the Office of Science and Technology Policy, during the administration of G.W. Bush, John Marburger, at the Executive Office of the President, in his Keynote Address at the 44th Robert H. Goddard Memorial Symposium March 15, 2006, said: '*The greatest value of the Moon lies neither in science nor in exploration, but in its material . . . I am talking about the possibility of extracting elements and minerals that can be processed into fuel or massive components of space apparatus. The production of oxygen in particular, the major component (by mass) of chemical rocket fuel, is potentially an important lunar industry!'*
[42] Schmitt, H.M., 2006, *Return to the Moon: Exploration, Enterprise, and Energy in the Human Settlement of Space*, Praxis Publishing, p. 335.
[43] Lewis, J.S., 1996, *Mining the Sky,* Addison-Wesley, ISBN 0–201–47959–1, p. 274.
[44] Sonter, M.J., 1998, *The Technical and Economic Feasibility of Mining the Near-Earth Asteroids,* 49th IAF Congress, Melbourne, Australia.
[45] Dyson, F., 1976, 'Can we control the carbon dioxide in the atmosphere?', *Energy* **2**, 287–291.
[46] Schiermeier, Q., 2006, 'Putting the carbon back, the hundred billion tons challenge', *Nature* **442**, 620–623.
[47] Jayaraman, K.S., 2007, 'India's carbon dioxide trap', *Nature* **445**, 350.
[48] Crutzen, P.J., 2006, Foreword to *Solar Variability and Planetary Climates* (ISSI Book Series No. 23), Calisesi Y. et al. (eds), Springer Publication and *Space Science Reviews* **125**, Issue 1–4, 1–3.
[49] Crutzen, P.J., 2006, 'Albedo enhancement by stratospheric sulfur injections: a contribution to resolve a policy dilemma', *Climatic Change* **77**, 211–220.
[50] Morton, O., 2007, 'Is this what it takes to save the world?', *Nature* **447**, 132–136.
[51] Angel, R., 2006, 'Feasibility of cooling the Earth with a cloud of small spacecraft near the inner Lagrange point (L1)', *Proceedings National Academy of Sciences* **103,** 17184–17189.

10 Managing the Planet's Future: The Crucial Role of Space

The Earth is blue like an orange.

Paul Eluard

10.1 Introduction

Throughout the previous chapters we have discussed the most potentially serious natural and anthropogenic hazards that confront our planet and challenge our capability to survive on it. They define the requirements for the tools to be developed in order to improve our understanding and to forecast and, whenever possible, control them and limit their consequences. Only recently have we been able to observe our planet 'from above' with aircraft and balloons. Space systems have been available for some decades. Depending on the altitude of the satellite, it is possible to get global, regional or local views with increasing resolution, from a few kilometers, down to meters or even centimeters in the case of military surveillance, as narrower field observations give information on local phenomena, on regional policies or even on individuals. Satellites are the only means we have to observe the Earth in its entirety, offering a nearly instantaneous snapshot on the physical status of the whole planet. The historical picture taken by the Apollo 8 astronauts in orbit around the Moon remains not only the symbol of our capability to reach and explore our mythic natural satellite, but has also given us the first global view of the planet on which we live, forcing us to reflect on its limits and its fragility.

In the past decades, the use of space has given us unprecedented views of the oceans, of the continents and of the poles across all geopolitical or national barriers. With them, we have been able to observe the degradation and recovery of the ozone layer since the 1970s and to separate the respective importance of anthropogenic and natural degradations. With satellites, we dispose of the most complete and most precise spectrum of information on the short- and long-term evolution of the Earth. It is now impossible for any nation to hide either the effects of these hazards or the way they cope with them to limit their consequences on the global environment and for avoiding their reccurrence.

The natural and human-induced disasters and the environmental factors affecting human health and well-being – such as cosmic hazards, volcanic

eruptions and climate change, but also weather forecasting and warning, the management of energy and mineral resources, of the water cycle and water resources, the evolution of terrestrial, coastal, and marine ecosystems, desertification and agriculture, biodiversity and soil conservation – all require continuously checking of the Earth's 'health', from the interior through the surface and the atmosphere, to the interplanetary medium. In that context, space tools definitely represent a strategic asset. Looking ahead to the next centuries, and in view of the unavoidable global character of our future, it is clear that the enormous potential of space observations, complemented with ground-based measurements, will make them more and more indispensable. They will certainly develop and expand, but their capacities must be coordinated on the world scale, while the permanence and continuity of their service must be preserved. We now describe the most important assets. A more detailed description can be found in reference [1].

10.2 The specific needs for space observations of the Earth

The Earth is a dynamic planet. For quite some time, for the sake of simplification, the Earth has been studied layer by layer from the interior to the upper atmosphere. The Earth could be compared to an onion. In effect, for quite some time the community of Earth scientists has been organized according to these layers. Of course, the Earth is not an onion but rather an orange as Paul Eluard, the French Poet, would say. All the components of the Earth System are in a permanent state of evolution and interaction: the interior, the hydrosphere (which includes the oceans, the aquifers and the dams), the cryosphere (which includes all the ice cover found in the polar caps, lakes and in glaciers and snow), the atmosphere (which includes the troposphere, the stratosphere, the ionosphere and the magnetosphere), and the biosphere. The timescales that characterize the evolution of these components vary from billions or millions of years down to tens of thousands, as well as shorter seasonal or diurnal variations. The accuracies required in term of time and spatial resolution (areas covered on the Earth) put severe constraints on the instruments. For example, the motion of plate tectonics is measured in centimeters per year and the rise of the sea level in millimeters per year. These precisions are now easily attained and well within the reach of present space-borne sensors. They will obviously improve in the future as better instruments are developed.

10.2.1 The Earth's interior

The interior of the Earth is probably the only part of our planet that is not directly affected (yet!) by anthropogenic activities. We could call it the 'astronomical Earth'. It is also referred to as the 'Solid Earth' or geosphere by analogy to the hydrosphere or to the atmosphere, even though it is not solid at all, since it contains a liquid iron and nickel core. It is the seat of natural planetary phenomena such as plate tectonics, volcanism and the dynamo that

generates the magnetic field. Its characteristics and properties fix limits to the whole Earth system. Their variations and long-term evolution have an influence on it. This requires continuous surveillance of what happens several thousand kilometers beneath our feet.

Although the interior of the Earth is not directly accessible for *in-situ* observations, our knowledge of these hidden parts has considerably improved owing to the availability of better ground-based studies, in particular paleo- and archeo-magnetism, as well as seismology measurements (see Chapter 4) and models. Studying all manifestations of phenomena that occur at the limits of the solid and the liquid core, upward to the mantle and to the crust, offers a powerful means of understanding the mechanisms that are driving the dynamics of the Earth's interior. Seismic waves and variations in the strength of gravity already provide a picture of the hot core, the rocky mantle and the crust. Magnetic measurements offer another means for understanding the Earth's interior. On the other hand, because the phenomena that have their seat in the interior are either global or cover a substantial area and volume of the planet, satellite-based measurements represent unique assets as they add their 'global' power to those performed from the ground. A tight correlation between data obtained by these two complementary approaches is essential for further progress. Because these phenomena have usually long time constants, the need for continuity is essential.

The generation of the magnetic field has been discussed in Chapter 2. The Earth's field is the superposition of the intrinsic field generated by the dynamo and of the fields generated in the upper layers, in particular in the ionosphere (see Section 10.3.8) and in the magnetosphere. Even though that second component is negligible, its effects can be important; for example, the magnetospheric field modulates the flux of cosmic rays that penetrate the atmosphere, possibly affecting the ozone layer and influencing the climate [2]. In the past 150 years, it has been observed that the axial dipole of the field has decayed by nearly 10%, a weakening that has been confirmed from evidence provided by the Oersted and Magsat space missions, and this phenomenon is characteristic of a pre-reversal situation (Chapter 2). Geographically, the present dipole decay can be attributed to changes occurring in the South Atlantic Ocean. Indeed, the field lines there deviate strongly from those of a pure dipole. This 'South Atlantic Anomaly', located west off the coast of Brazil at latitudes between 35 and 60 degrees, most likely results from the eccentric displacement of the center of the magnetic field from the geographical center of the Earth by some 450 km, and has a net effect on satellite detectors and communications as it lowers the radiation belts by several hundred kilometers, exposing spacecraft and their equipments to high doses of radiation (Figure 10.1). Interestingly, the altitude of the anomaly is not stable and is at present moving closer to the ground. It also drifts to the west at a speed of about 0.3 degree per year, very close to the rotation differential between the Earth's core and its surface. This secular variation directly reflects the fluid flow in the outer core and is one of the characteristics of the Earth dynamo. Continuous monitoring of the South

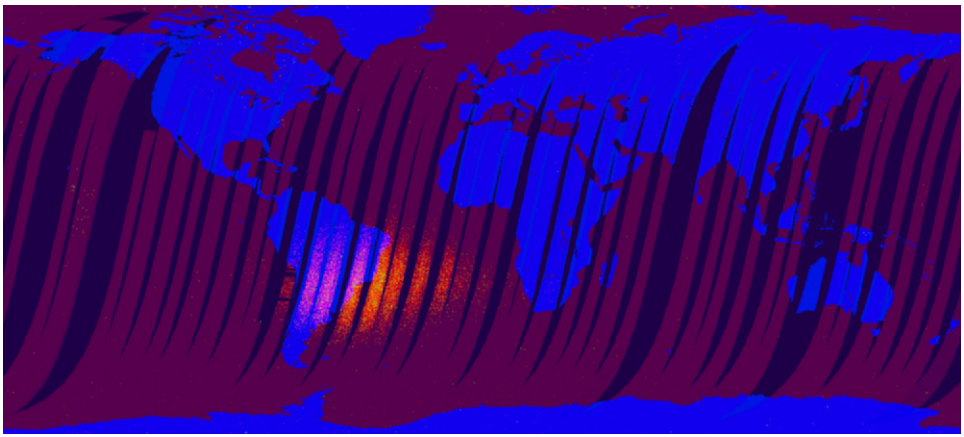


Figure 10.1 The South Atlantic Anomaly as 'seen' by the Multi-angle Imaging SpectroRadiometer (MISR) instrument on board NASA's Terra spacecraft. The MISR cameras, designed to detect visible light, are also sensitive to energetic protons at high altitudes. With the cover closed, no light hits the detectors and only the background levels of protons stand out. The long 'S' structures correspond to the successive orbits of the satellite. (*Credit*: NASA/GSFC/JPL, MISR Science Team.)

Atlantic Anomaly and of the fluctuations of the field may cast some light on the mechanisms of the field's generation and on the dynamics and properties of the inner core of the Earth – two problems that have not yet been solved.

Variations or modifications of the Earth's magnetic field might induce substantial modifications of the magnetosphere, weakening its capacity to shield the Earth against the penetration of cosmic and solar wind particles (Chapters 2 and 3). Continuous space-borne monitoring in low Earth orbit and the establishment of models of the field will be very important for predicting these potential hazards within the space environment. Artificial satellites provide a unique tool for monitoring the global dipolar Earth's magnetic field, the behavior of the magnetosphere and how it reacts to solar wind activity. The Swarm mission of ESA (see Box 10.1) aims at exactly addressing such needs [1].

Of more immediate concern are the hazards that result from the tectonic motions, volcanic eruptions and earthquakes. Here also, the combination of ground-based networks of stations and of space systems prove to be unique for monitoring the deformation of the ground, fault ruptures and volcanic eruptions, eventually forecasting their catastrophic occurrence. Following the displacements of the plates with a precision of a few millimeters allows us to reconstruct the seismic history of a given area, which is the first step toward forecasting. Space techniques allow a better surveillance and lessen the dependence on ground-based instruments. For example, space-borne radars can peer through vegetation and follow with a precision of millimeters how plates are moving and how the strain is building up before earthquakes and eruptions. It is now possible to measure the slow deformation in between

Box 10.1 The Swarm mission

The objective of ESA's Swarm mission is to provide a survey of the geomagnetic field and its temporal evolution, and gain new insights into improving our knowledge of the Earth's interior and climate. Swarm consists of a constellation of three satellites in three different polar orbits between 400 and 550 km altitude. High-precision and high-resolution measurements of the strength and direction of the magnetic field will be provided by each satellite. GPS receivers, an accelerometer and an electric field instrument will provide supplementary information for studying the interaction of the magnetic field with other physical quantities describing the Earth system. For example, Swarm could provide independent data on ocean circulation.

A new generation of magnetometers will enable measurements to be taken over different regions of the Earth simultaneously. Swarm will also provide monitoring of the time-variability aspects of the geomagnetic field, which is a great improvement on the current method of extrapolation based on statistics and ground observations. The geomagnetic field models resulting from Swarm will further our understanding of the atmospheric processes related to climate and weather and will also have practical applications in many different areas, such as space weather and radiation hazards.

earthquakes and obtain the rate of strain accumulation in a given region. With the growth of the population, new centers of habitation are occupying areas of greater risk and there is a clear need to systematically monitor the motions and displacement of the ground in view of mitigating the consequences of the related hazards. The deformations of the solid Earth also induce changes in the global sea level, modifying the boundaries between land and water. The timescales for such effects might be very long, but the hazards may occur in a very short time. They range between hundreds of millions of years for convection to just a few seconds for earthquakes. In the following section we shall describe how and why space geodesy and altimetry, Global Positioning Systems and radar interferometry represent the most promising and powerful tools for observing and monitoring the rising or sinking motions of the ground and the displacements of large masses and large portions of the Earth's surface, including the subsidence of the ground [3].

10.2.2 Water: the hydrosphere and the cryosphere

Water is responsible for the extreme complexity of the Earth system when compared, for example, with Venus or Mars. It is the medium through which the most important chemical reactions for life take place and the most precious and indispensable resource for life. We discussed in Chapter 8 the situation of water in the context of increasing demands, and the need to preserve its quality. We

have also seen that there is no scarcity of water on the Earth, therefore, it is of great importance to know where and in what form the water is distributed: oceans, rivers, glaciers, or polar caps. The global water cycle – the transport and distribution of large amounts of water, associated with its constant phase changes between solid, liquid and gaseous states – is one of the most important features of the Earth system. Today, the lack of global data is the major constraint on the development of water resources and improvement of water management. Every year satellite observations are becoming more essential in that respect as they characterize water-related processes on both the global and the regional scales. Here also, the combination of space-based data with high-resolution *in-situ* data is mandatory.

The largest share of water resources is in the seas, and oceans are the central heating system of the planet through the thermohaline circulation mechanism (Chapter 5). Oceanic circulation plays a crucial role in the recycling of carbon dioxide and may induce cooler or warmer climates in a way that is not yet fully understood. The salinity of water is modifying the oceanic circulation and changing the currents, and it must be monitored. In addition, the total volume of liquid water is strongly temperature dependent through two additive effects: melting of ice and thermal expansion. Sea-level rise is one of the most worrisome problems for populations living along the coastal areas and is a problem that requires a complete system approach: it involves interactions between the atmosphere, the hydrosphere, the cryosphere and the biosphere. Furthermore, mass exchanges between these components and changes in the mass distribution in the Earth's interior, must be known precisely in order to quantify their relative contributions to sea-level rise [1].

Water vapor in the atmosphere also affects the weather and the climate, though scientists are only beginning to understand how these complex mechanisms work. It is one of the most efficient greenhouse gases. An increase in water vapor in the atmosphere may increase the amount of clouds. Another typical effect of water is its contribution to the capacity of Earth to reflect incoming sunlight not only through these clouds but also through snow and ice which have reflective broad band albedos varying between 60 and 90%.

The polar caps concentrate some 90% of the entire planet's ice and 80% of its fresh water reserves. Polar regions are an integral part of the climate system; through their high albedo, they contribute in an important way to the natural cooling of the planet. But, most threatening, the melting of the ice cap, as discussed in Chapter 6, would contribute substantially to sea level rise. Gravity changes measured between 2002 and 2006 by the gravimetry satellite GRACE, which is sensitive to mass changes, indicate that Greenland lost between 212 and 284 km^3 of ice, which is higher and faster by a factor of 2 than many previous estimates [4]. The melting of that ice has resulted in an increase of 0.5 mm of the global sea level. If it were to melt entirely, the global sea level would rise by about 7 meters and tens of thousands of kilometers of coastal zones would disappear. But this is nothing compared to Antarctica. We now know from recent satellite observations that the western part of Antarctica is providing one-fifth of the

present rise in global sea level, representing 0.16 mm a year [5]. If Antarctica were to melt entirely, its ice content would raise the global sea level by 70 meters! However, the temperatures in Antarctica are so low that even with increases of a few degrees, they would mostly remain below the melting point of ice.

Satellite radar data show that snow seasonally covers up to 30% of the land surface. Global warming will raise the snowline up 150 meters for every 1°C increase, with rain precipitation occurring at those altitudes that previously received snowfall. Permafrost occupies a quarter of the exposed land area in the northern hemisphere. It plays an important regulating role in the water cycle and the exchange of gases between the land and the atmosphere, in particular methane, as has been detected through space observations [1]. Continuous monitoring with spatially resolved observations of the cryosphere is therefore essential to assess its sensitivity to climatic variations and to study the effects on water and gas exchange with the atmosphere.

While thermal expansion is a less obvious process than ice melting (mainly because you cannot see it happening), the IPCC predicts that thermal expansion will be the main component of expected sea-level rises over the 21st century (Figure 10.2). Recent observations by the Topex–Poseidon and Jason 1 satellite

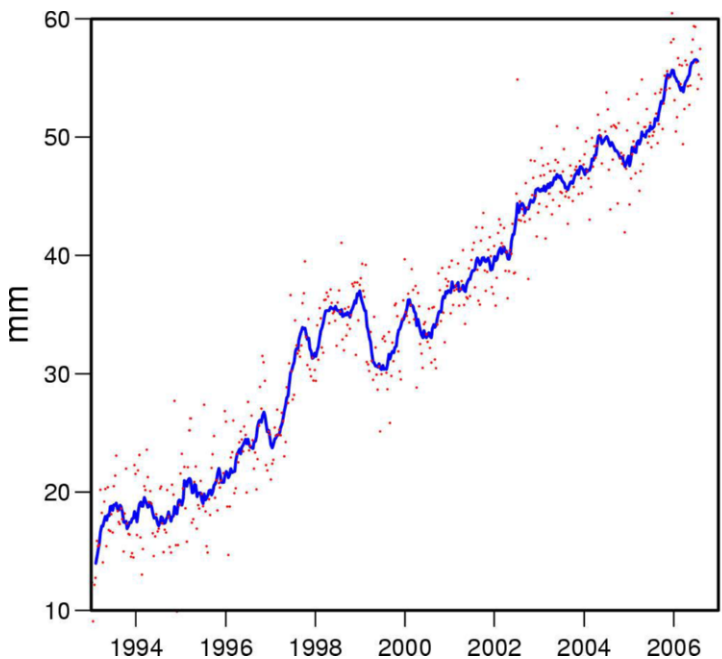


Figure 10.2 Satellite altimetry measurements of sea level rise measured by the Topex-Poseidon and Jason 1 satellite altimetry (data averaged over 65°N and 65°S between 1993 and 2006. The red dots are raw 10-day sea level data. The blue curve corresponds to a 60-day smoothing of the raw 10-day data. The annual seasonal periodic variations have been removed. (*Credit*: LEGOS-CNES-IRD, Courtesy A. Cazenave.)

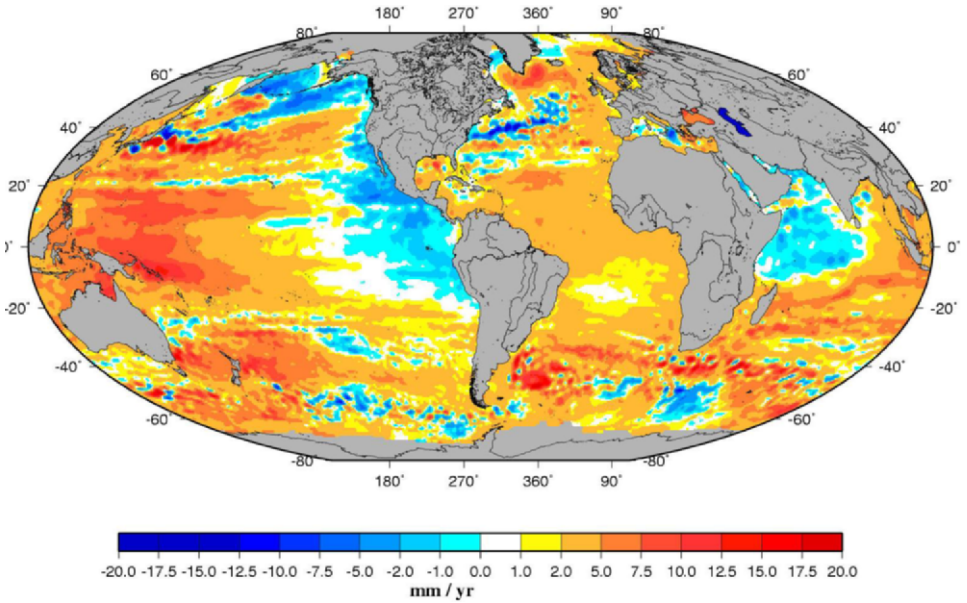


Figure 10.3 Regional distribution of sea-level trends measured by the Topex–Jason 1 altimetry satellites between January 1993 and June 2006. Variations from region to region are essentially due to the effect of thermal dilatation [6]. (*Credit*: LEGOS–CNES–IRD, Courtesy A. Cazenave.)

indicate that it is in the western Pacific and eastern Indian oceans that sea-level rise shows the greatest magnitude, an effect due essentially to thermal dilatation [6] (Figure 10.3). In many places, a best estimate of 35 cm as proposed by the IPCC, within a range of 22 to 50 cm, would see entire beaches being washed away, together with a significant chunk of the coastline. For people living on low-lying islands such as Tuvalu, Kiribati or the Maldives, where the highest point is only 2–3 meters above current sea levels, the upper limit of 50 cm could see significant portions of their islands being washed away by erosion or covered by water. Even if they remain above the sea, many island nations will have their supplies of drinking water reduced because sea water will invade their fresh water aquifers. These islands have sizeable populations, but these are insignificant compared to the tens of millions of people living in the low-level coastal areas of southern Asia: Pakistan, India, Sri Lanka, Bangladesh and Burma. A mean sea-level rise of between 15 and 38 cm is projected by the mid-21st century along India's coast. Added to this, a projected increase in intensity of tropical cyclones (Chapter 4) would significantly enhance the vulnerability of populations living in these cyclone-prone coastal regions.

These space data are completed by *in-situ* measurements such as performed by the international Argo system which uses 3,000 profiling floats distributed in all the oceans and can dive at depths of 2,000 meters undersea and transmit their

data on temperature and salinity regularly via satellites providing over 100,000 profiles each year – which is more than 20 times greater than the annual similar measurements made by scientific and merchant vessels, and from areas where these ships cannot go because of ice coverage.

10.2.3 The atmosphere

More than any other part of the Earth system, the atmosphere is shared by everybody independent of any territorial division. Like the stars above, the air is a common good: we live in it under one bar of oxygen and nitrogen plus a few other constituents. We breathe it and use it in our daily lives, for ground and air transportation. But this vital element is very fragile. Even though the atmosphere extends to altitudes of several 100 km through the troposphere and the stratosphere (Figure 10.4), its scale height [7] is only 8.5 km, and 90% of its mass is concentrated below an altitude of 16 km, or just 0.0025 of the Earth's radius, which gives an idea of its thinness. The composition of the atmosphere can be modified naturally through outgassing and volcanic activity, by living organisms or as a result of anthropogenic industrial activities, and to a lesser extent by meteoritic bombardment. Any modification in the chemical content of

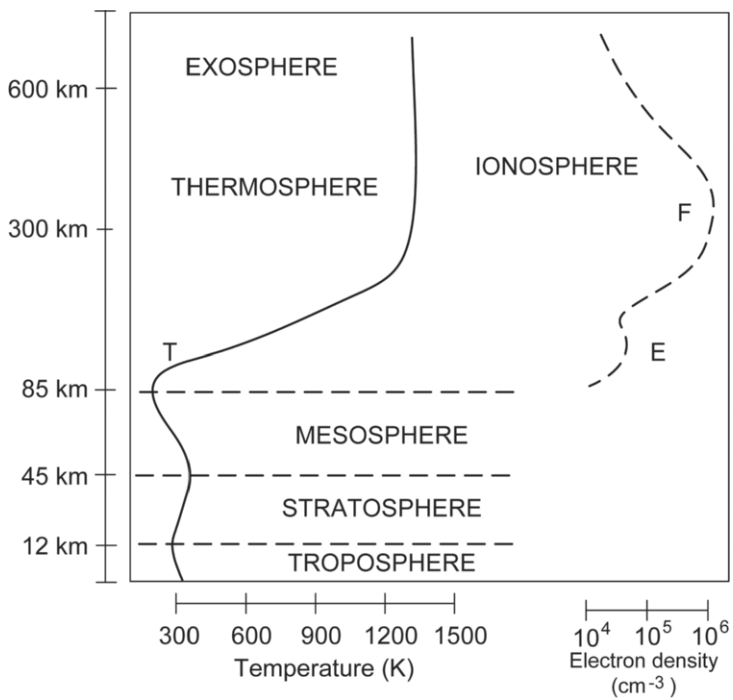


Figure 10.4 Distribution with altitude above the Earth's surface of atmospheric temperature for the various layers. The figure also shows the E and F layers of the ionosphere and their electron densities.

greenhouse gases (GHG) may have important consequences on the climate and, with rising temperatures, on the height of the oceans (Figure 10.2 and 10.3). Continuous monitoring of that chemical composition and its modifications is an essential requirement for a proper approach to safeguard the planet's future. All gases emitted by natural and anthropogenic sources have no alternative but to go into the atmosphere, which is naturally very sensitive to changes of chemical content.

The atmosphere is a thermodynamic as well as a dynamic system, pervaded by winds which change and shear with altitude, by vertical currents and cyclones. It is also subject to powerful electrical phenomena such as thunderstorms and is directly influenced by cosmic and solar radiation. Together with the magnetosphere it acts as a protective shield against lethal radiations, in particular solar ultraviolet, thanks to the ozone layer in the stratosphere. It interacts directly with water through evaporation, clouds and aerosols which modify the albedo and the portion of solar radiation that is absorbed in the various layers and the portion that reaches the ground. It also interacts with the biosphere through respiration, cattle flatulence and agriculture. Understanding the various mechanisms at play there, and the way the atmosphere reacts to natural and anthropogenic perturbations, is therefore indispensable for ensuring the living conditions of all life on Earth.

The oxygen we breathe represents 21% of the atmosphere per volume. Carbon dioxide, in a proportion of just a few hundred parts per million (ppm), helps to maintain the sea and the surface at an average temperature of 15°C through the greenhouse effect (Chapter 5). The total number of carbon atoms on the Earth is more or less fixed, if we exclude the contribution from the – still existing – meteoritic bombardment. As explained in Chapter 2, carbon is recycled through the ocean, the atmosphere and the biosphere and its concentration in the various parts of the Earth's system varies with time (Chapter 5). Only about half of the anthropogenic emissions sent into the atmosphere stay there. The rest is probably absorbed in the oceans and in the land, with around 14% of worldwide carbon stored in permafrost soils and sediments. Tundra wetlands are considered to be major contributors to the global carbon balance, and are anticipated to be highly sensitive to climate change: if they were to suddenly outgas, the resulting global warming effects would be much more dramatic than at present. This is why it is so crucial to follow the complete carbon cycle and observe the exchanges between the land, the ocean surface and the atmosphere. The observations must not only address the monitoring of carbon dioxide in the atmosphere but also surface monitoring, including forest and vegetation cover, tundra, fires, biomass, humidity and sea and land photosynthesis. Two projects, the Orbiting Carbon Observatory (OCO) prepared by NASA and the Greenhouse Gas Observing Satellite (GOSAT) in Japan, are intended to measure carbon dioxide (see Section 10.3.6).

Another life-protecting atmospheric constituent is ozone. In the stratosphere, ultraviolet sunlight breaks oxygen molecules and causes them to temporarily combine with an oxygen atom forming O_3, the ozone molecule (Chapter 3).

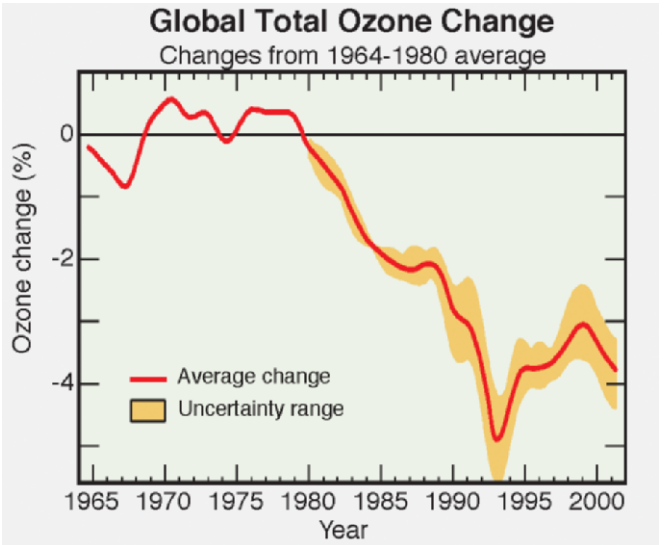


Figure 10.5 Global total ozone changes between 1964 and 2002 as measured by space-borne instruments and compared with the average corresponding to the period of 1964 to 1980. The global total ozone content has decreased by an average of a few percent in the last two decades. Between 1980 and 2000, the largest decreases occurred following the Pinatubo eruption in 1991. In the 1997 to 2001 period global ozone was reduced by about 3% from the 1964–1980 average. (*Source*: UN Environment Program [8].)

Ozone helps to filter most of the UV-B band and more energetic ultraviolet radiation from the Sun. Because organic components of life are strong absorbers of lethal UV radiation, ozone plays a unique role in the preservation of life. Its existence is indeed one of the factors that permitted life to exist on land (aquatic organisms, including all known early life forms, are shielded by water). Known health hazards induced by UV radiation include increased mutation rate, skin cancer and cataracts, depression of the immune system, impaired crop and tree growth, and the death of plankton. Each 1% drop in ozone is thought to increase human skin cancer rates by 4–6%. Ozone is very sensitive to anthropogenic pollutants like the chlorofluorocarbons (CFCs), in particular Freon (a refrigerant) and halocarbons which are released by fire extinguishers, which destroy the molecule. Permanent monitoring of ozone concentration is therefore another essential measurement. Such measurements have in fact been performed nearly continuously since the early 1960s (Figure 10.5). They evidence both the degradation of the ozone layer by human activities as well as volcanism, in particular the strong effect of the Pinatubo eruption, and should soon be able to observe the effect of the countermeasures imposed by the Montreal Protocol on the limitation of GHG usage. The largest decreases have occurred at the highest latitudes in both hemispheres because of the large winter/spring depletion in

polar regions. The losses in the southern hemisphere are greater than those in the northern hemisphere because of the greater losses that occur each year in the Antarctic stratosphere. Long-term changes in the tropics are much smaller because reactive halogen gases are not abundant in the tropical lower stratosphere [8].

Aerosols contribute to the thermal balance of the Earth and should be properly monitored. They are made of liquid or solid particles in suspension in the atmosphere. Some are produced directly by the dispersion of particles emitted from the ground, while others are the product of transformations of atmospheric substances forming particles. Several million tons of aerosols are emitted daily, coming from a large variety of sources both natural (volcanic, biologic, desertic, marine) and anthropogenic (burning, industrial dusts, agriculture) – Figure 10.6. They are mostly found in the troposphere where their time of residence can reach several days. Being strongly influenced by rain precipitation, their concentration is fairly inhomogeneous at the regional scale as opposed to the GHG which tend to be more global.

They play a fundamental role in influencing the air quality and the climate, and their interactions with clouds is very important, but yet poorly understood because of their complexity [9]. They counteract the warming effect of the GHG because they intercept sunlight, resulting in less energy reaching the Earth's surface, hence cooling! But some also absorb light, increasing the warming of the atmosphere and giving a potential reduction in cloudiness. Small aerosol droplets may also increase the lifetime of clouds and thereby the Earth's albedo,

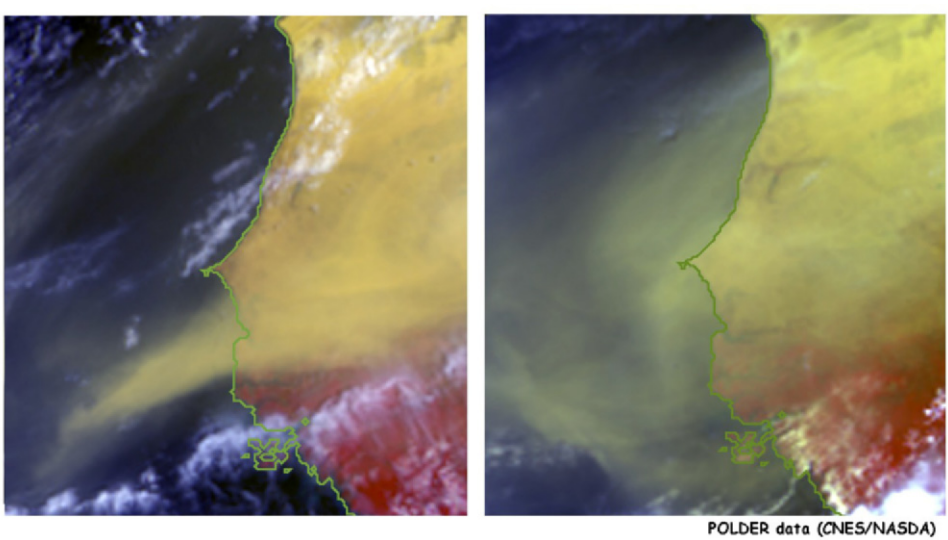


Figure 10.6 Desertic aerosols spreading over the west coast of Africa. *Left*: 29 May 1997; *right*: one day later and showing the evolution of the dust clouds observed by the POLDER 1/ADEOS 1 satellite instrument. (*Credit*: POLDER–PARASOL.)

amplifying their cooling effect. The knowledge of these complex and indirect actions is essential to properly assess the counteracting effects of GHG and aerosols on global warming. Future progress will come from both space-based observations and *in-situ* local measurements by means of high-altitude balloons and lidar ranging.

10.2.4 The biosphere

The biosphere, or ecosphere, ranges from about 10 km above the ground into the atmosphere down to the deepest ocean floor, including most of the lower atmosphere, the hydrosphere and the upper lithosphere. There, all living organisms are found interacting with one another and with their environment. The estimated number of identified living species is about $\sim$1.75 million, but it is likely that the total number is above 30 million and may be as high as 100 million, to be found mostly in yet unexplored zones of tropical forests and jungles. Therefore, this number is likely to increase as more species are identified, while, at the same time, biodiversity is currently being lost across the globe at a rate unprecedented in human times, due to the disappearance of many species that are victims of anthropogenic activities. It is therefore understandable that the mass of the biosphere cannot be estimated very accurately, and to answer key environmental, agricultural and health questions, biodiversity scientists are obliged to base their predictive models on incomplete data.

Biodiversity is necessary for the sustained delivery of the goods and services that are essential for human well-being, as well as for the maintenance of life on Earth in general. Our food, part of our energy, fibers, the control of pests and diseases and the discovery of novel natural products, such as pharmaceuticals, all rely on biodiversity. Anthropogenic activities and the growing world population are rapidly transforming the ecosphere, by putting increasing pressure on, in particular, the land, the hydrosphere and the atmosphere. Growing cities are definitely eliminating large parts of fertile lands that can no longer be used to provide the necessary food. Vegetation is the key component of the biosphere. Phytoplankton, in particular, accounts for the majority of the biomass in the oceans, and has a greater effect on our planet's climate through the recycling of carbon than any other living species – including all the world's forests (Figure 10.7). Vegetation represents our ultimate source of subsistence. It is the source of the oxygen we breathe and of the amino acids from which all animals and humans build up their own proteins (see Box 10.2). Monitoring the evolution of the biomass and more specifically, of the global vegetation and phytoplankton reserves also appears to be essential to ensure our survival from now on!

The biosphere interacts with the water cycle, the carbon cycle, the energy cycle and the climate. These interactions are very dynamic and, as we have seen many times, are very complex and not well understood because the solution of such a complex problem requires a large variety of data and scientific analysis that are not available today with the proper degree of accuracy. Satellite remote-sensing data provide a macroscopic view of the ecosphere status and the



Figure 10.7 Summer marine phytoplankton bloom filling much of the Baltic Sea as observed on this image captured by the MERIS instrument on board ENVISAT on 13 July 2005. (*Credit*: ESA.)

continuity of observations is crucial if we are to follow what appears to be a rapidly evolving situation. In order to understand and evaluate biodiversity, and accurately predict the consequences of further loss, many sources of observations must ideally be pooled together. Most of them are, and will continue to be, made *in situ*. However, a coherent global system of observations would greatly improve our capacity of analysis and prediction (Figure 10.8).

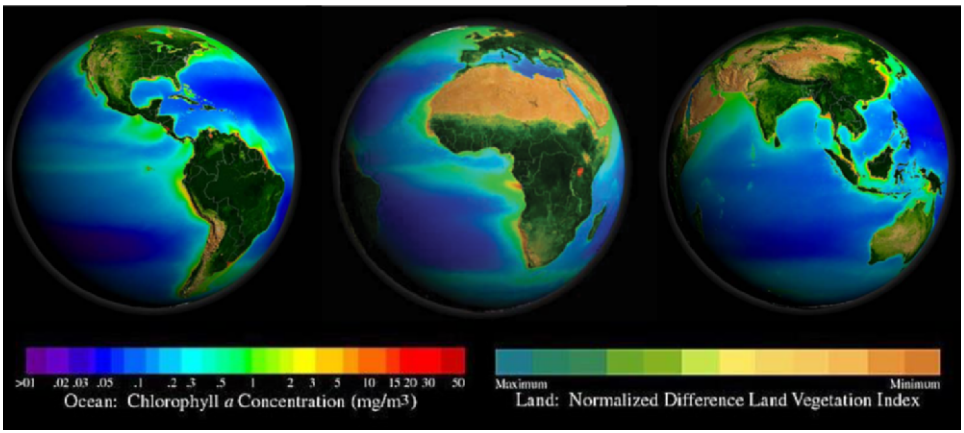


Figure 10.8 The global biosphere derived from the SeaWiFS satellite images. (*Credit*: NASA/Goddard Space Flight Center Scientific Visualization Studio.)

Box 10.2 The Normalized Difference Vegetation Index

NDVI is calculated from the amount of visible (VIS) and near-infrared (NIR) light reflected by vegetation. Nearly all satellite Vegetation Indices employ the following difference formula to quantify the density of plant growth on the Earth:

$$NDVI = (NIR - VIS)/(NIR + VIS)$$

Healthy vegetation absorbs most of the visible light that hits it, and reflects a large portion of the near-infrared light and has a high index. By contrast, unhealthy or sparse vegetation reflects more visible light and less near-infrared light and has a low index.

10.3 The tools and methods of space

From what precedes, we see that understanding the Earth's system, its dynamics and how it evolves, requires a continuous monitoring of the key parameters of each of its components as well as the analysis of the complex relations that interconnect them. The tools, both scientific and technical, must permit altogether the monitoring of the Earth's surface, be it solid or liquid, of its shape and motions, and of its temperature and vegetation cover. They must also be able to measure the composition of gases in the atmosphere, the concentration and circulation of aerosols and their temperature. They should support studying and forecasting the climate, if possible at a similar degree of reliability as we manage today when studying and monitoring the weather. They should

allow the forecasting of natural as well as of anthropogenic hazards. They must offer the capabilities of continuously measuring the incoming solar radiation, and ultra-violet flux. Satellites are capable today of making these measurements with the required degree of accuracy on both a global and a local scale. Wherever possible they must be completed by *in-situ* measurements for both cross-checking and calibration of the data and for reaching parts of the system that cannot be observed from orbit, such as the ocean's subsurface layers. That implies that they must be operated and integrated into a global system. An exhaustive description of the tools in operation, or under development, in the various organizations and space agencies of the world is given in reference [10].

10.3.1 The best orbits for Earth observation

Orbits are optimized according to the goal of a mission. High-resolution imaging instruments are usually placed on Low Earth Orbits (LEOs) at altitudes of a few hundred kilometers. On the other hand, orbits located far away from the Earth – such as those centered around the Lagrange point L_1, located between the Sun and the Earth, where the gravitational attraction of each body exactly cancels – offers the possibility of observing the Earth globally as it moves along its orbit around the Sun and slowly rotates on its axis. This orbit has been considered for observations of a global character such as the thermal and radiative balance of the Earth [11]. If, in the future, the Moon is exploited as a scientific platform, it is quite likely that there will be instruments on its surface for global or low-resolution Earth observations.

The period of the orbit (the time it takes for a satellite to pass over the same longitude or latitude point) is a function of the altitude of the orbit and varies from approximately 90 minutes for the satellites in LEO to several hours or days for higher orbits. The inclination of the orbit plane relative to the equator is an important parameter. An equatorial orbit (whose plane corresponds to the Earth's equatorial plane) will of course favor the observation of equatorial and tropical zones, while a satellite in polar orbit (whose plane is perpendicular to the equatorial plane) will permit the observations of the whole globe and, in particular, of the poles with a repetition time that depends on the period. The altitude of a polar orbit can be adjusted in such a way that its period is an exact fraction of a day: in a Sun-synchronous orbit, the satellite passes over the same zone of the Earth at roughly the same local time each day. This makes communications and various forms of data collection very convenient. For example, a satellite in a Sun-synchronous orbit could observe the air pollution over Paris or Beijing at noon every day. A typical orbit particularly suited for meteorological satellites is the geostationary equatorial orbit at $\sim$35,800 km with a period of 24 hours, so that the satellite viewed from the Earth is a fixed point in the sky. Such orbits are particularly useful for observing the variations of the same zone at any time of the day. The French SPOT-4 satellite in geostationary orbit has a 101.5-minute period with an exact repetitivity of 26 days and a quasi-repetitivity of $\sim$5 days. Reference [10] also contains the respective orbits of the various Earth observation satellites.

10.3.2 Geodesy and altimetry satellites: measuring the shapes of the Earth

Gravimetry satellites

Gravity is the fundamental force that influences many of the processes of the Earth system. If the Earth were a perfect sphere, the force of gravity would have a constant value everywhere on its surface and equal to 9.81 m/s^2. In reality, the Earth is not a perfect sphere. Its shape closely resembles an ellipsoid with the equatorial radius being about 21 km greater than the polar radius. This phenomenon is called the Earth's oblateness and is due to the effect of the centrifugal forces induced by the Earth's rotation. Consequently, the Earth's gravity varies from 9.78 to 9.83 m/s^2 from the equator to the poles. The gravity field and the shape of the Earth define a surface called the *geoid*. This is the surface that the Earth would have if it were entirely covered with oceans in the absence of winds, currents and other disturbing forces. Departures from a perfect ellipsoid are represented by the geoid elevation above or below the ellipsoid. The geoid can be as low as 106 meters below the ellipsoid or as high as 85 meters over areas of several thousands kilometers. The geoid is a reference surface whose precise shape allows the determination of the irregularities and time variations in the distribution of mass which induce variations in the gravity field. The different materials that make up the layers of the Earth mantle and the crust are not homogeneously distributed, and their thickness is not regular. For example, the crust beneath the oceans is thinner and denser than the continental crust. Figure 10.9 shows the geoid for three areas of the Earth, as established by the GRACE gravimetry space mission and measured in mGals, a unit equal to 10^{-5} m/s^2, or about one millionth of the Earth's acceleration of gravity [12].

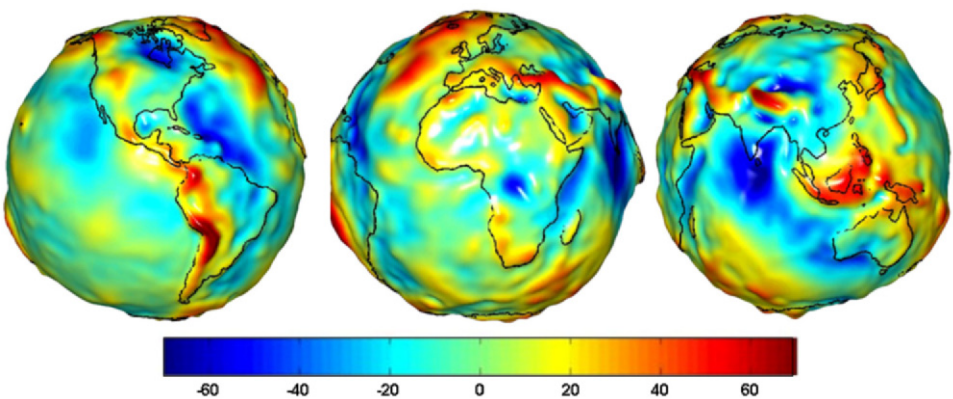


Figure 10.9 The reference geoid as observed by GRACE. The Earth's oblateness has been removed. The gravity field anomalies are measured in mGals on the color scale and gravity variations have been artificially exaggerated for the purpose of better viewing. (*Credit*: University of Texas Center for Space Research and NASA.)

The long timescale variations in the gravity field represent the effect of large-scale convective cells in the Earth's mantle. The medium to short timescale changes are mostly due to variations in the distribution of the water content as it cycles between the atmosphere, oceans, continents, glaciers and polar ice caps. They range from tens to hundreds of mGals. The deviation of local sea level from the geoid can therefore be closely linked to the ocean circulation whose changes are a consequence of changes in atmospheric forcing, primarily caused by surface wind stress and heat and fresh water flux. Precisely measuring these variations is therefore crucial for understanding the internal structure of the Earth, the dynamics of the Earth system and the climate. This is the task of gravimetry – and altimetry – satellites and of the GPS described in the next section. Contrary to weather or other remote-sensing satellites where imagery is playing the main role, geodesy, altimetry and positioning satellites are performing their measurements through the precise knowledge of their positions and of their motions relative to the Earth's surface, or to themselves in the case of multi-satellite systems.

Gravimetry satellites are rare. Only three have contributed substantially to the field. CHAMP from Germany was launched in 2000 and exploits satellite-to-satellite tracking between an orbit at 400 km and the US GPS satellites at about 20,000 km, and has led to impressive improvements in gravity field models for features of up to a few thousand kilometers, in particular for polar regions which are usually difficult to access.

GRACE, launched in 2002, is made of two satellites, co-orbiting at near polar inclinations at 400 km and separated by about 220 km. As the gravity beneath them shifts with the local density of the Earth, the separation between the two satellites varies and is accurately measured by means of the GPS. This technique leads to an improvement of several orders of magnitude in the gravity measurements and allows much better resolution of the broad-to-finer scale features of the gravitational field over both land and sea. It has the capability of mapping monthly changes for features of 600–1,000 km. Early in its mission, scientists faced problems in exploiting GRACE data to their expected resolution because of the difficulties encountered in representing the various coefficients used to parameterize the gravity field [13]. That situation derives from the very principle of the GRACE concept which is not omni-directional but rather privileges the dimension of the two satellites axis. The best accuracy presently achieved is estimated to be less than 1.8 cm RMS for smoothed features of 750 km, and 2.4 cm for those of 500 km [14]. GRACE has monitored the variation of water reservoirs such as in the Amazon basin (Figure 10.10).

ESA is responsible for developing the Gravity Field and Steady-State Ocean Circulation Explorer, GOCE, which will take advantage of a low-altitude orbit at 250 km – which is more sensitive to the gravitational signal – to establish global and regional models of the Earth's gravity field with 1–2 mGal precision, and a geoid with 1 cm accuracy over about 100 km of spatial resolution. At such a low altitude, the atmosphere is relatively dense and the lifetime of the satellite is limited to about 20 months. GOCE is a one-satellite mission as opposed to

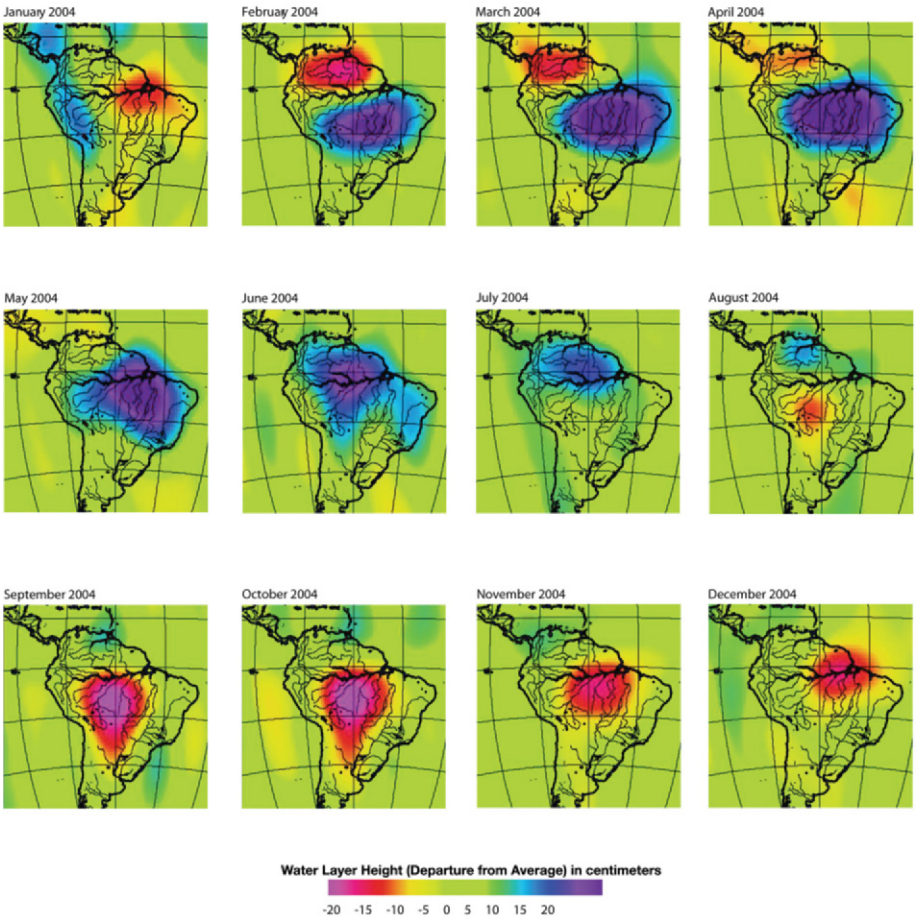


Figure 10.10 The amount of water flowing through the Amazon basin varies from month to month, and can be monitored from space by looking at how it alters the Earth's gravity field. This series of images was produced using data from GRACE and shows monthly changes relative to a 3-year average over the Amazon basin and neighboring regions. Oranges, reds and pinks show where gravity is lower than average; greens, blues and purples show where gravity is higher than average. The Amazon has distinct rainy and dry seasons, and the seasons show up clearly in the monthly maps. Notice also that the smaller Orinoco basin to the north of the Amazon has a distinctly different seasonal pattern. (*Credit*: University of Texas Center for Space Research.)

GRACE and it will not, in principle, suffer from the problems of its predecessor. The gravity gradiometer measurements will be completed by satellite-to-satellite tracking relative to 12 GPS satellites. GOCE will use a sun-synchronous, near circular orbit with a 96.5 degree inclination. GOCE aims at advancing

research in the fields of steady-state ocean circulation and the physics of the Earth's interior [15].

Altimetry satellites

Whereas gravimetry satellites use the alterations of their orbits to measure the distribution of mass concentrations on the Earth as they pass over them, altimetry satellites directly measure the altitude of their orbit relative to the Earth's surface features. Altimetry satellites use radars or lidars, which probe the Earth's surface by sending radar or laser pulses. The waves are echoed back to the satellite from a limited area of the Earth's surface corresponding to the beam illumination. The altitude or distance is directly deduced from the time it takes the wave to travel from the satellite and back. Precise positioning of the satellite, using the GPS or an equivalent system, with respect to a reference like the geoid, allows distances to be measured relative to that reference. In this way the altitudes of continents and their mountains, of glaciers and of waves on the ocean can be properly monitored.

If the Earth possessed a thin atmosphere, say like Mars, altimetry could be carried out with lasers beaming from orbit to the planet's surface, as laser light can be easily transmitted through a thin atmosphere. The Earth's atmosphere, however, is thick and has clouds, and visible light lasers, which cannot operate through clouds, work only during day time. By contrast, radars operating at frequencies of the order of 1 to 10 gigahertz (corresponding to wavelengths of 30 to 3 cm), whose light is unabsorbed along its way from the satellite down to the ground and vice versa, allow day-and-night observations. Since radar waves can pass though clouds, they also offer an all-weather service. However, as these waves have longer wavelengths than those of laser light, the final resolution on the ground is usually coarser than for laser altimeters and of a few hundred meters. The motions of the ground, i.e. tectonic plates, can be measured with radar altimetry with an accuracy of 1–2 mm/year. Radar waves can also penetrate underground, but the depth of penetration is a function of the soil humidity: the smaller the deeper!

Satellite altimetry has proved to be irreplaceable in determining the undulations of the Earth gravity field on scales of a few kilometers as the sea surface irregularities reproduce the undulations of the bottom of the sea, which can reach tens of meters – i.e. a factor 10–100 larger than the ocean's perturbations – and provide unique information on the mechanics of tectonic plates and the dynamics of the Earth's mantle [16]. Radar altimetry has been in operation since the mid-1970s on board US as well as Canadian and European satellites, in particular the couple ERS 1, 2 and ENVISAT from ESA, and the US–French Topex–Poseidon and its successor Jason 1. In the mid-1980s, the American satellite, GEOSAT, had established a high-resolution altimetric map of the ocean geoid for the US Department of Defense. These data remained classified for strategic reasons for some 10 years. In 1994, ERS 1 was moved to an orbit that permitted high-precision cartography of the ocean surface with a resolution of a few kilometers. The combination of these two sets of data have led

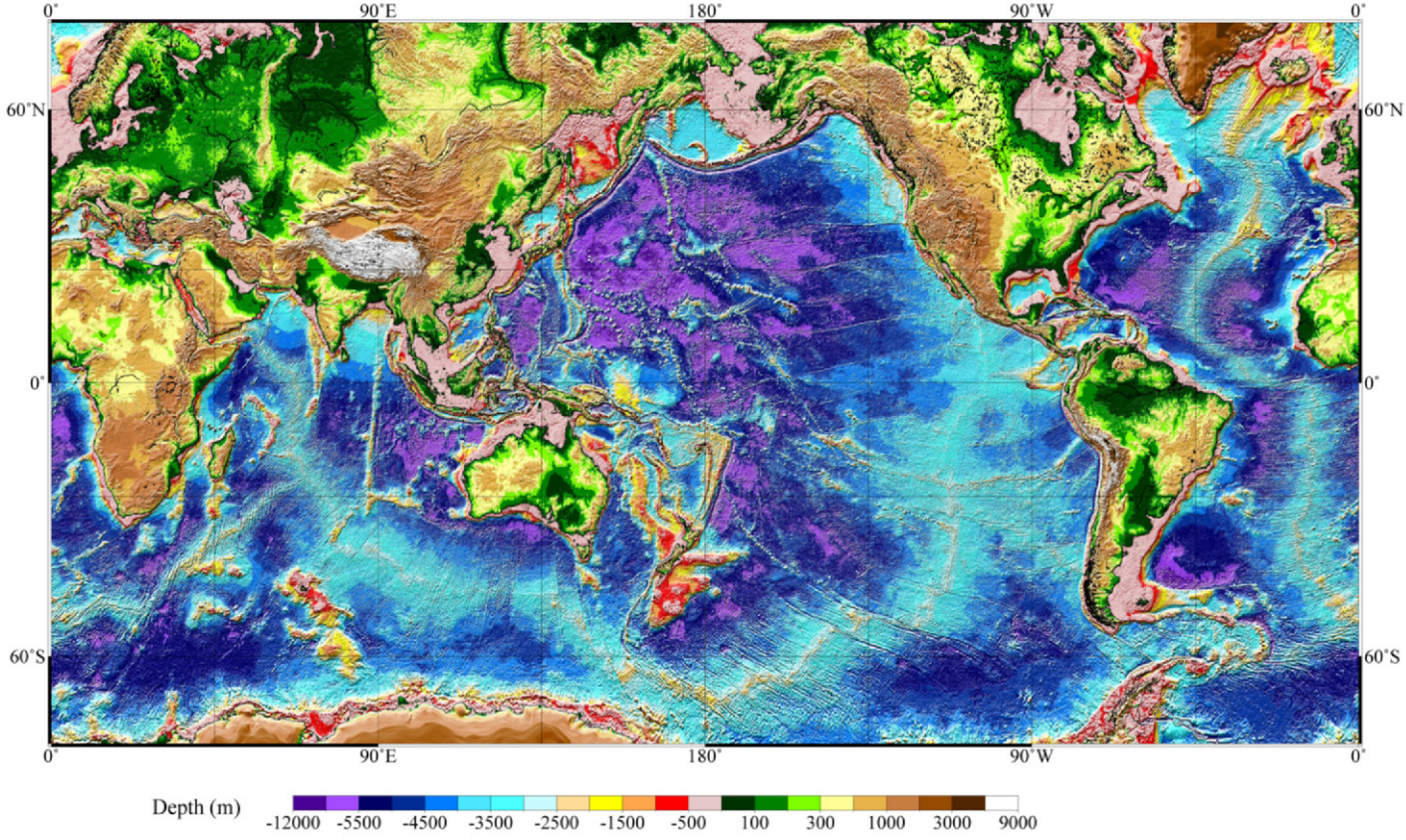


Figure 10.11 Map of the ocean floor as established through space altimetry combining the data from ESA's mission ERS 1 and the American GEOSAT. (*Credit*: LEGOS and CNES.)

to the establishment of high-resolution maps of the topography of the ocean floor, evidencing the great complexity of plate fractures, alignment of submarine volcanoes and large quantities of fossil structures, the testifying to intense tectonic activity in the past (Figure 10.11).

With a precision of a fraction of 1 mm, radar altimetry allows the measurement of the elevation of sea level due to global warming as evidenced on Figure 10.2 which shows an increase of 3.0 ± 0.4 mm/year of the sea level on average over the period 1993–2006. This is due to both the thermal dilatation of the water over the last 10 years, counting for 1.5 mm/year, and the melting of polar caps plus water from continental reservoirs, rivers and ices which contribute another equal amount [6]. Radar altimetry has also proved to be invaluable for the observation of the polar ice caps, as demonstrated by the ERS 1 and 2 and ENVISAT missions. The technique is so powerful and the measurements so important that fully dedicated missions are now envisaged to fulfill that goal, such as NASA's ICESAT and ESA's CryoSat [5], a three-year radar altimetry mission, that will determine variations in the thickness of the Earth's continental ice sheets and marine ice cover, and test the prediction of thinning

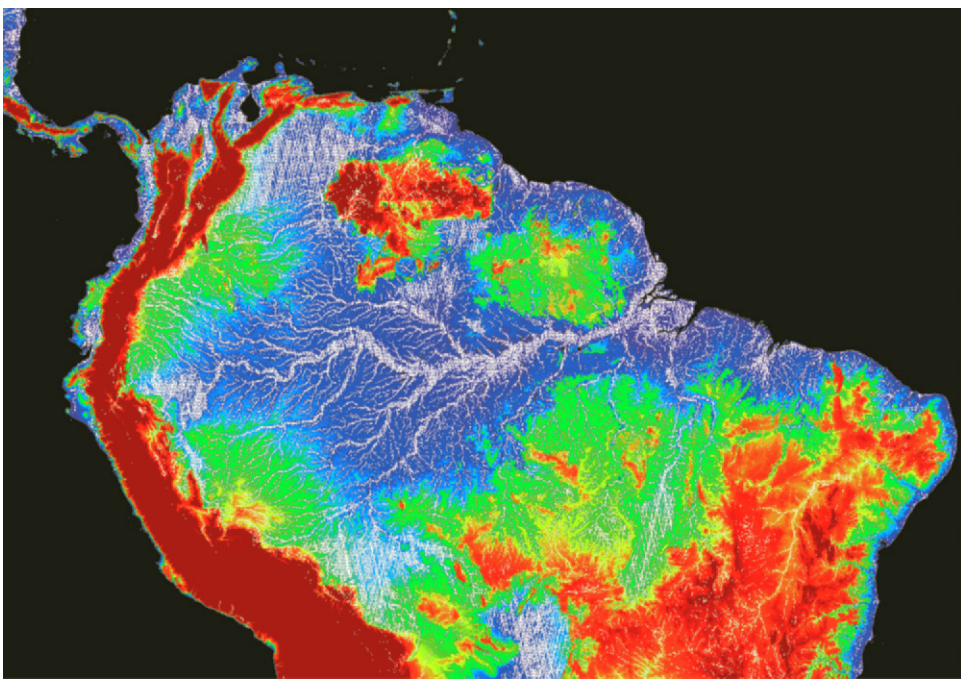


Figure 10.12 Radar altimetry data provides researchers with the means to monitor global river and lake levels. This image shows the Amazon river basin observed with the Radar Altimeter on ERS 1, including 'wet' radar echoes from rivers, lakes and swamps. The straight linear structures on the pictures are artifacts produced by the radar imaging technique. (*Credit*: ESA.)

Arctic ice due to global warming, with an accuracy of about 0.12 cm/year over areas of more than 10^7 km^2. It has also been demonstrated with ERS that echoes from inland water surface are clearly discernible and convertible to river or lake levels [17]. The majority of the world's river systems can now be monitored (Figure 10.12) and continuity in the observations makes it possible to survey the river heights and hydrology of the whole planet.

10.3.3 Global Positioning Systems

Global Positioning Systems (GPSs) have been in operation since the mid-1980s both in the USA and in USSR/Russia (GLONASS), under the control of the military for clear strategic purposes. Europe is now developing a new independent system called Galileo. All these system operate according to the same principle. Timing signals emitted though extremely accurate atomic clocks are transmitted at known times from a number of satellites at known locations. By measuring the times at which these signals are received, the respective distances of the various emitters and receivers can be deduced.

For the system to work with a positioning accuracy of a few centimeters, the precision of the time measurements must be 0.1 of a billionth of a second. Besides the uncertainty of the clock itself, owing to its constant movement and variable height relative to the Earth-centered inertial reference frame, the clocks on the satellites are affected by the relativity theory. General relativity predicts that the atomic clocks at the GPS orbital altitudes will tick more rapidly, by about 45.9 microseconds per day, because they are in a weaker gravitational field than atomic clocks on Earth's surface. Inversely, Special relativity predicts that atomic clocks moving at GPS orbital speeds will tick more slowly than stationary ground clocks by about 7.2 microseconds per day. When combined, the discrepancy is about 38 microseconds per day, a difference of 4.465 parts in 10^{10}. It is remarkable that the fantastic development of the GPS and their impressive number of applications, rest on a direct application of Einstein's theory. This single example illustrates the fundamental role of science in the development of modern civilization and of society. In the future, this role will continue to expand and is a key to securing our survivability on this planet. We address this fundamental issue in the next chapter.

The space segment of the present American system was based originally on a constellation of 24 satellites orbiting at an altitude of approximately 20,200 km (orbital radius of 26,600 km) and distributed equally among six circular orbital planes centered on the Earth, of approximately 55 degrees inclination relative to Earth's equator. Each satellite passes over the same location on Earth once each day. The orbits are arranged so that at least six satellites are always within line of sight from almost everywhere on the Earth's surface. The accuracy of positioning varies from a few tens of meters to a few centimeters. Much greater accuracies in the range of a few millimeters can now be achieved through differential corrections of the positions of different GPS satellites and by tracking the phase of the carrier signal. The use of radio frequencies makes the GPSs all-time all-weather operating systems. As of September 2007, the US GPS rests on 31 actively

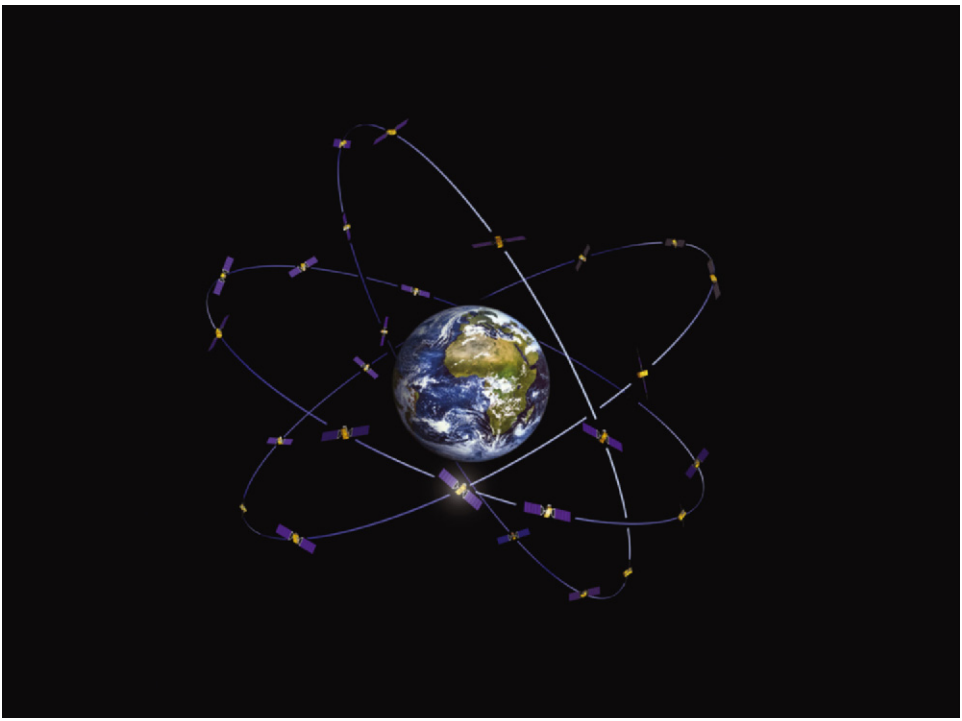


Figure 10.13 Configuration of the 30 Galileo satellites in their orbits around the Earth. (*Source*: ESA.)

broadcasting satellites which improve the precision by providing redundant measurements, as just explained. The European Galileo system plans a set of 30 satellites placed on three circular orbits at an altitude of 23,616 km. Typically, any single point on the Earth is always in view of 8 to 12 satellites of the constellation (Figure 10.13). It will provide an accuracy of a few meters for civilian use and a few centimeters for commercial use.

The applications of the GPS are indeed incredibly numerous, making them truly indispensable for civilian and scientific purposes. They range from precise measurements of the rotation of the Earth (therefore the length of the day) and its orientation in space, through the motions of the crust, of plates and microplates, and earthquakes. In the latter case, motions and sliding of a few millimeters can be measured, providing within a few minutes important information on the magnitude of the earthquake and of the probability that it might later induce a tsunami. Similarly, GPS stations located on the flanks of volcanoes can provide a warning that the volcano is undergoing structural changes [18] (see Chapter 4). GPS stations are also providing an essential support to all geodesy and altimetry satellites. They are also used for the monitoring of the evolution of ice caps and for the studies of the atmosphere using the occultation of the carrier signal through the different atmospheric layers. GPS

radio waves are particularly well suited for measuring the absorption by atmospheric water and for determining the vertical column density, the profiles of pressure and temperature, and the structure of the troposphere through limb sounding [19]. A very clever use of the GPS allows sea surface winds to be measured by observing the extent to which the signals are scattered after reflection by the waves.

At altitudes above 100 km the ionosphere can also be studied using the GPS signals as the speed of radio waves is affected according to the density of electrons. It turns out that GPS signal distortion caused by the ionosphere varies as the radio frequency of the signal changes, while atmospheric distortions remain constant. This technique does not only allow the distortion of the GPS signal to be measured but can also distinguish how much of the observed distortion is caused by the atmosphere and how much is caused by the ionosphere. The Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) mission, comprising six micro-satellites under the joint responsibilities of the USA and Taiwan, is exploiting a technique that also seems to be very promising in the prediction of cyclones and weather forecasts.

10.3.4. Synthetic Aperture Radars

Besides their utilization as altimeters, radars can be used also in the Synthetic Aperture Radar (SAR) mode to build images (see Figure 10.14 and Box 10.3). SARs are operating day and night and in all weather conditions. They produce pictures

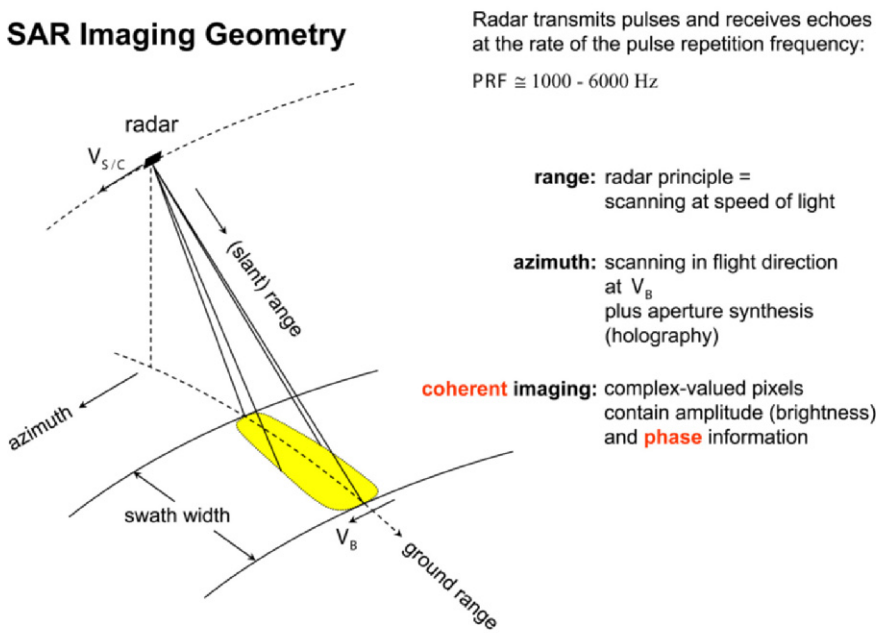


Figure 10.14 The principle of SAR imaging geometry. (*Courtesy*: R. Bamler, DLR.)

Box 10.3 Measurement principle of Synthetic Aperture Radars

The geometry of operation of a SAR is shown on Figure 10.14 and a brief description of the principle can be found in references [20] and [21].

The antenna beam of side-looking radars is directed perpendicular to the flight path and illuminates a swath parallel to the satellite's ground track. Owing to the motion of the satellite, each target element is illuminated by the beam for a certain period of time called the 'integration time'. The echo signals received during this period are added coherently. The radar light is emitted in pulses with a frequency known as the Pulse Repetition Frequency, which usually reaches several thousand hertz. The two coordinates most commonly used are the 'range' or distance directly perpendicular to the antenna of the radar, and the 'azimuth' or distance along the flight path. To build a two-dimensional image, one must discriminate the signals of the radar wave energy scattered by the ground features. In the range direction, this is done by precisely timing the echoes. In the azimuth direction this is done by tracking changes in frequency caused by the Doppler effect. The image is built by combining the echoes of many pulses, creating a synthetic receiving aperture which mimics the performance of a much larger antenna whose dimension is the distance the radar antenna moved during the pulses. SAR resolution can reach 30 meters in range and 5 meters in azimuth. The most recent TerraSAR from Germany, operating at a frequency of 9.65 gigahertz, reaches a resolution of 1 meter over an area of 5×10 km^2, which is matched only by optical imagery.

of a scene in both two and three dimensions. SARs have been flying since the late 1970s, inspired by the success of the short-lived NASA SEASAT mission. They are able to indicate changes with time in soil and ocean situations and are well suited for differentiating between waterlogged and dry land (Figure 10.15). They prove to be particularly useful to study the evolution of droughts and flooded areas, such as in the case of Sahel nomads. SAR images over oceans are providing an indication of the state of the surface. If the surface were flat like a mirror, the image would appear dark except in the direction of the geometrically reflected beam – a very improbable situation for the receiving antenna on board the satellite. By contrast, an agitated surface is scattering the radar waves over a broad angle and a fraction of them are reflected in the direction of the satellite giving the image a gray color whose different tones are indicative of such ocean features as waves and currents. For example, SARs have proved to be excellent in following oil spills as the roughness of the water is strongly diminished by the presence of oil which lowers the scattering of radar waves and make the spills appear darker than the sea itself (Figure 10.16). SARs are also proving very powerful for the study of polar and ice-covered areas in general (see Figure 5.3). ESA's Cryosat mission is equipped with such a capability.

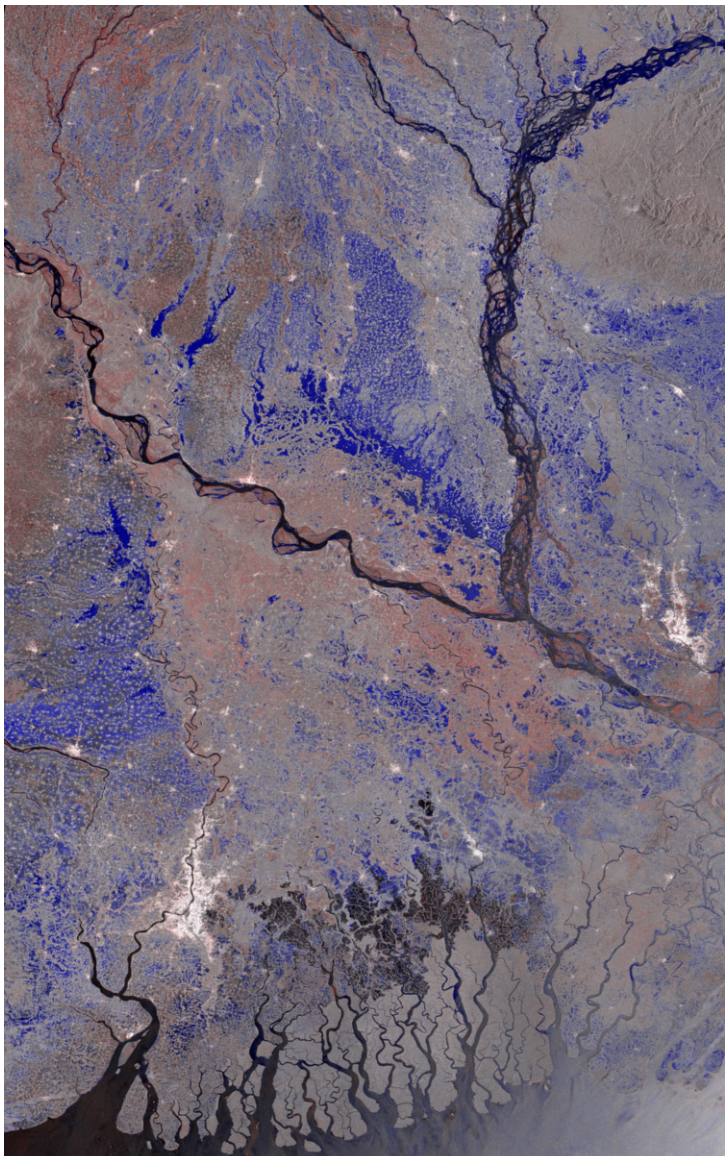


Figure 10.15 Flooding in Bangladesh and parts of India brought on by two weeks of persistent rain as observed by the Advanced Synthetic Aperture Radar (ASAR) on board ENVISAT. The image is a composite of two: one acquired on 26 July 2007 and another on 12 April 2007. Areas in black and white denote no change, while areas outlined in blue are potentially flooded spots. Areas in red-brown may also indicate flooding, but could also be related to agricultural practices. The bright white area on the bottom left of the image is Calcutta. Dhaka, the capital of Bangladesh, is visible as the bright white area in the center right. The mouth of the Ganges is visible in the center, and the Brahmaputra river is seen in dark. (*Credit*: ESA.)

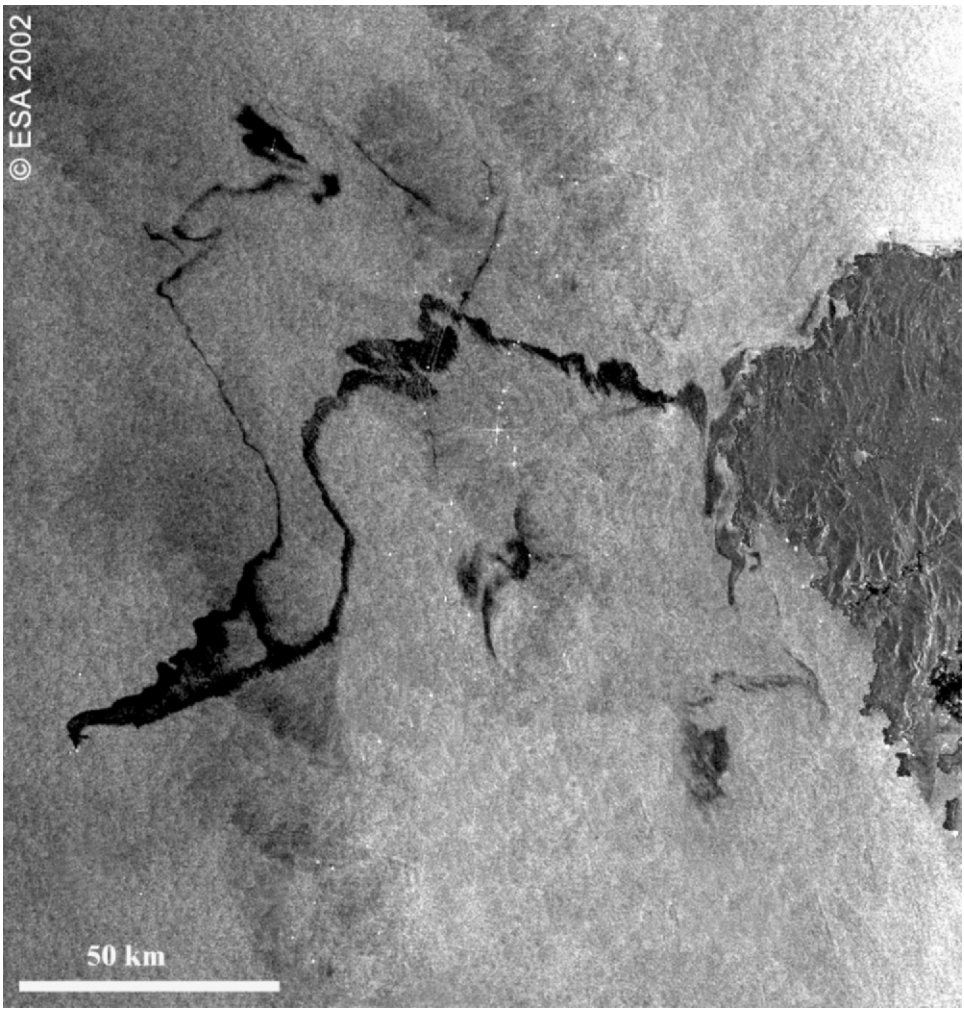


Figure 10.16 This ENVISAT ASAR image shows the oil spill originating from the stricken *Prestige* tanker, lying 100 km off the Spanish coast. The image was acquired in emergency on 17 November 2002. It covers an area 400 × 300 km. (*Credit*: ESA.)

Synthetic Aperture Radar interferometry, or InSAR, represents one of the most innovative applications of the SAR. This technique, first exploited on board aircraft, requires several satellite passes over the same area in order to obtain a suitable pair of radar images acquired from, as near as possible, the same point in space at different times. The orbital cross-track separation constitutes the interferometer baseline. The technique consists in mathematically combining the signals echoed from the same area at two different moments or from two satellites. This allows the tracking of small changes in the Earth's moving surface which would be otherwise undetectable. The ESA ERS 1 and ERS 2 tandem has

operated in this mode for several years. Their orbits were adjusted in order to cover the same track on the ground. For each pixel corresponding to the same area on the ground, an 'interferogram' is built, which reveals the changes in the distance separating the ground and the radar antenna on board the spacecraft, resulting from modifications of the Earth's surface due to, for example, tectonic motions or the swelling of volcanic magma chambers. In this way, one can derive more information than just a two-dimensional image, in particular the deformation of the surface. These maps provide an unsurpassed spatial sampling density of $\sim$100 pixels/km^2, a precision of 1 cm, and an observation cadence of 1 pass per month.

Geophysical applications of radar interferometry have exploded in the early 1990s as the technique is ideal for recording movements in the crust, perturbations in the atmosphere, dielectric modifications in the soil, and relief in the topography. A complete description of these capabilities can be found in reference [20]. The most important geophysical applications are those related to landslides, earthquakes, volcanoes and glaciers.

For example, knowing the orbit parameters, a Digital Elevation Model (DEM) of the surface can be extracted from the phase interferometric data. Such models are used to establish the elevation of the relief and evaluate the distribution of ground mass over areas which may be inaccessible on the ground (Figure 10.17). Because of the rugged relief or of vegetation cover, the average accuracy of a DEM can reach a few meters depending upon the terrain's topography and the surface-cover coherence. DEMs are important tools in many Earth-science disciplines. They greatly improve flood forecasting through modeling the watershed hydraulics, and helping to determine water availability for irrigation, power production, industrial and agricultural production from local slope information. In glaciology, they have already demonstrated their potential for determining the topography over ice sheets and glaciers, and for deriving ice motions. Combined with spatial-analysis models, they can be used to identify and simulate viewing perspectives for land-use planning, soil management and flood mitigation (Figure 10.18).

Differential interferometry also allows the quantification of small topographic changes and the assessment of surface dislocation and subsidence due to earthquakes or volcanoes with a precision of 1 cm or better. This is demonstrated on Figure 10.19 [22] in the case of the 26 December 2003 earthquake in Iran near the Bam region where the dislocation can be measured before and after the event with two different passes of ENVISAT over the area. Similarly, and probably more useful in view of forecasting potential eruptions, is the continuous monitoring of the modification of the flanks of volcanoes as done presently with Etna (Figure 10.20) [23, 24]. The combination of radar and other instruments such as the GPS has provided particularly important information on the swelling by 18 cm between July 2004 and the end of 2006, or an average of 7 cm per year of the Yellowstone 60-km-long caldera in the center of the US National Park. The caldera is the result of a gigantic eruption that occurred 642,000 years ago. Twelve GPS stations and ENVISAT ASAR have combined their power to detect the

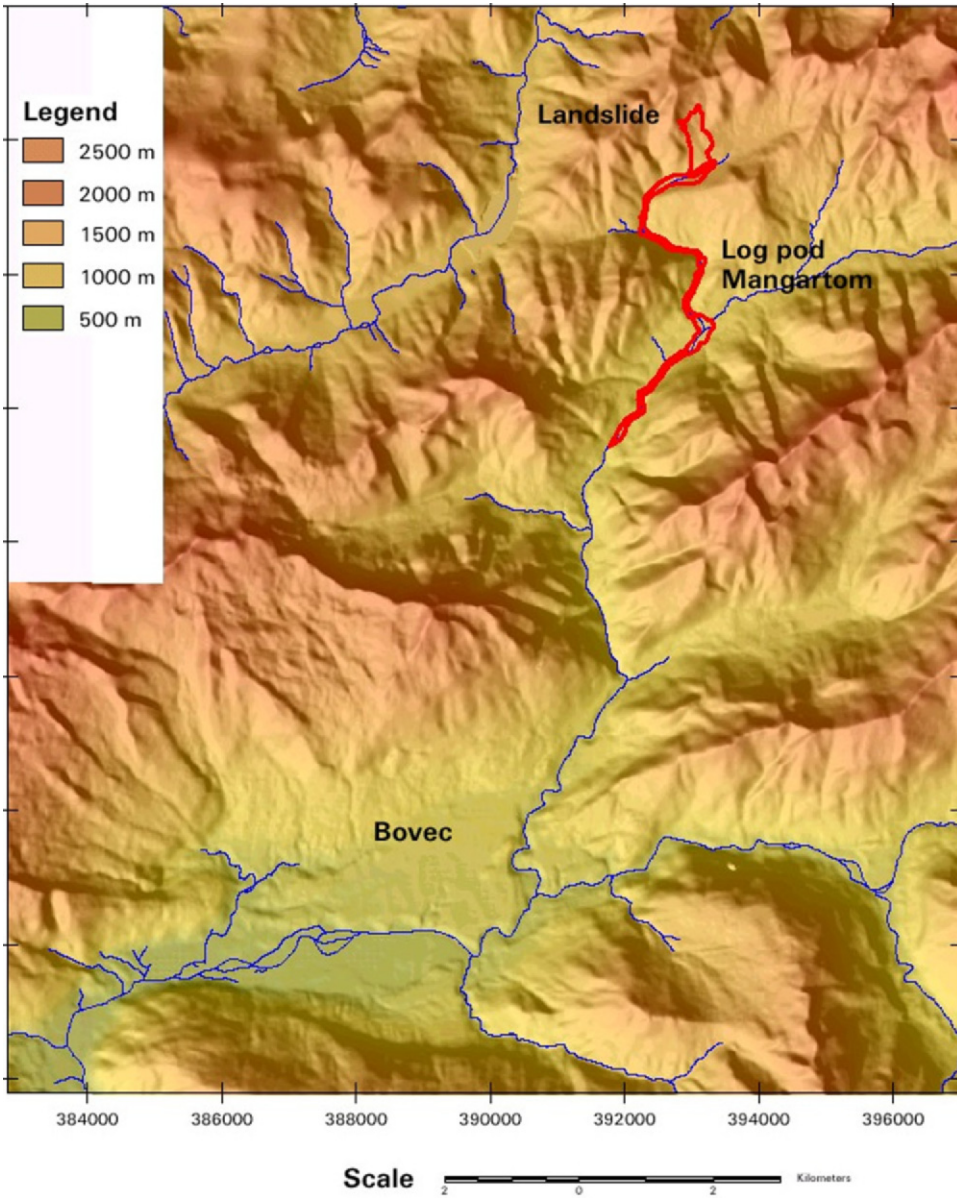


Figure 10.17 Digital Elevation Model of the 15 November 2000 Log Pod Mangartom landslide in Slovenia established with ERS 1 and 2 images. Landslide and damage areas are shown as vector overlays. (*Credit*: Scientific Research Center of Slovenian Academy of Sciences and Arts and ESA.)

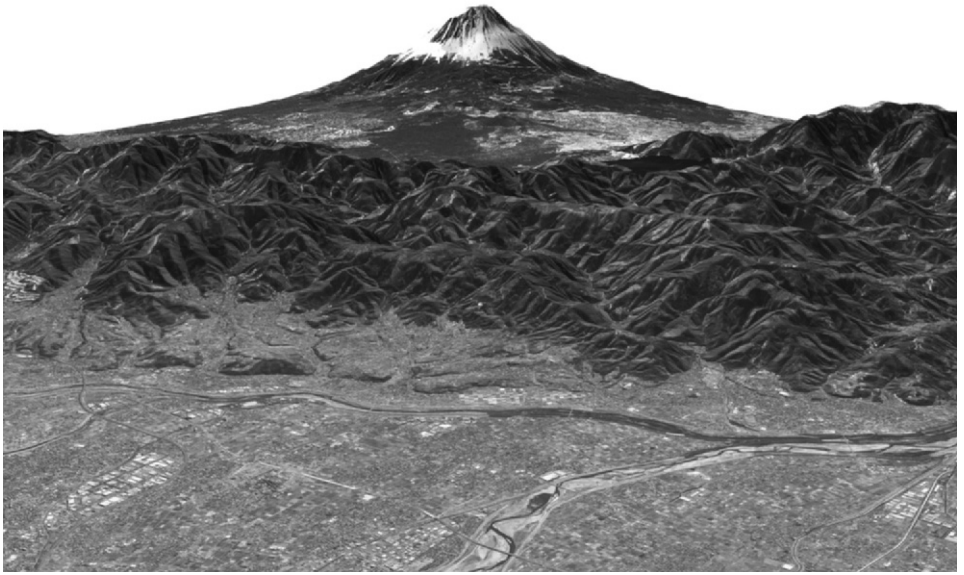


Fig.10.18 The city of Tokyo 'climbing' the slopes of Mount Fuji as observed by the Synthetic Aperture Radar on board the Japanese Advanced Land Observing Satellite (ALOS) 'Daichi'. (*Credit*: JAXA.)

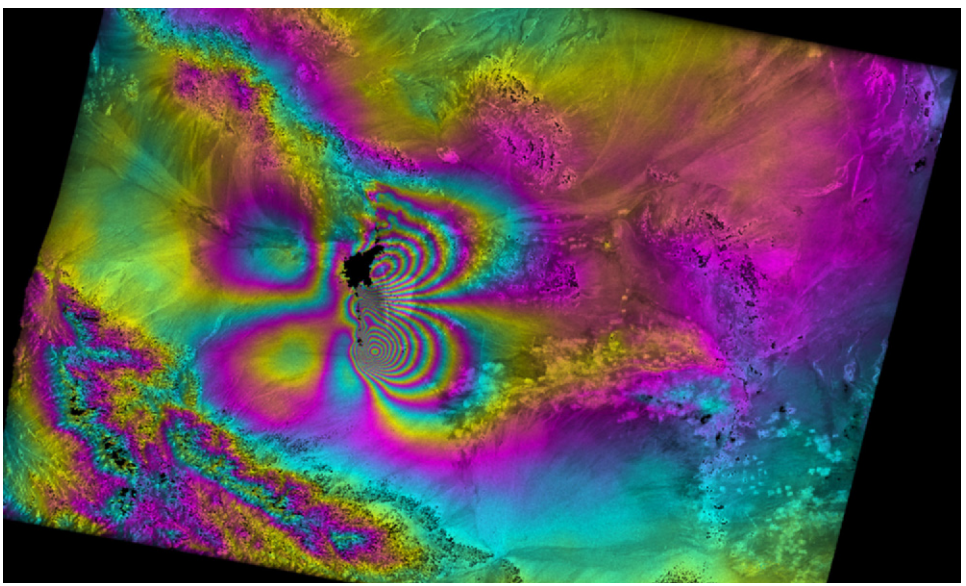


Figure 10.19 Interferometric map of the 26 December 2003 earthquake in Iran near the Bam region. This interferogram was built from images taken by ENVISAT 23 days before and 47 days after the earthquake. Each color fringe represents a displacement of the ground of 3 cm along the direction of the antenna's length. (*Credit*: Y. Fialko, University of California, San Diego, and ESA [22].)

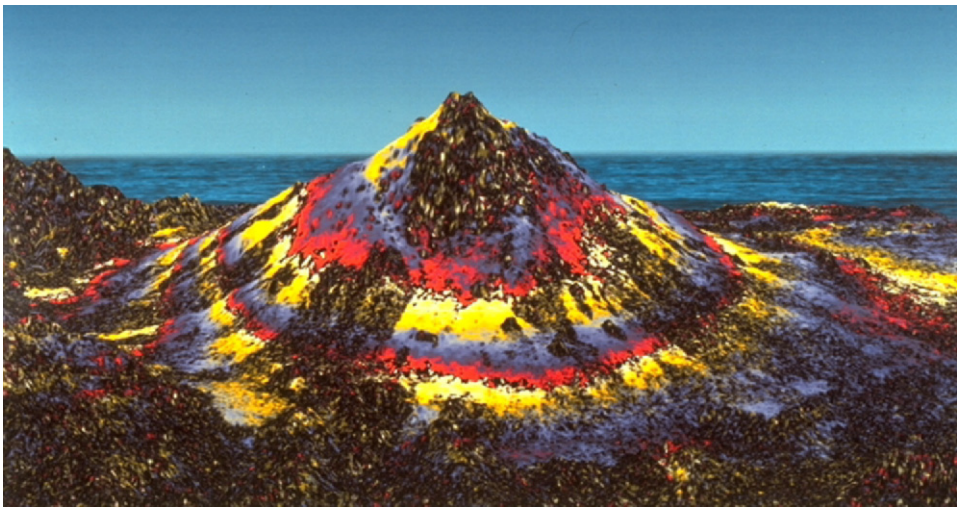


Figure 10.20 This three-dimensional view of Etna (3,300 meters) has been overlaid with SAR fringes from the ERS 1 and 2 tandem. Each color fringe corresponds to one half radar wavelength, or 28 mm. The precision of the displacements can reach 1 mm. This image shows the unswelling of the volcano consequent to the decreased internal pressure in the magma chamber following the 1993 eruption cycle. The succession of four fringes corresponds to a total amplitude of about 12 cm. The sky and sea background is an artifact introduced in the picture for reality enhancement purposes [23]. The three-dimension presentation has been built with a ground numerical model derived from the SPOT optical satellite. (*Credit*: CNES and IPGP.)

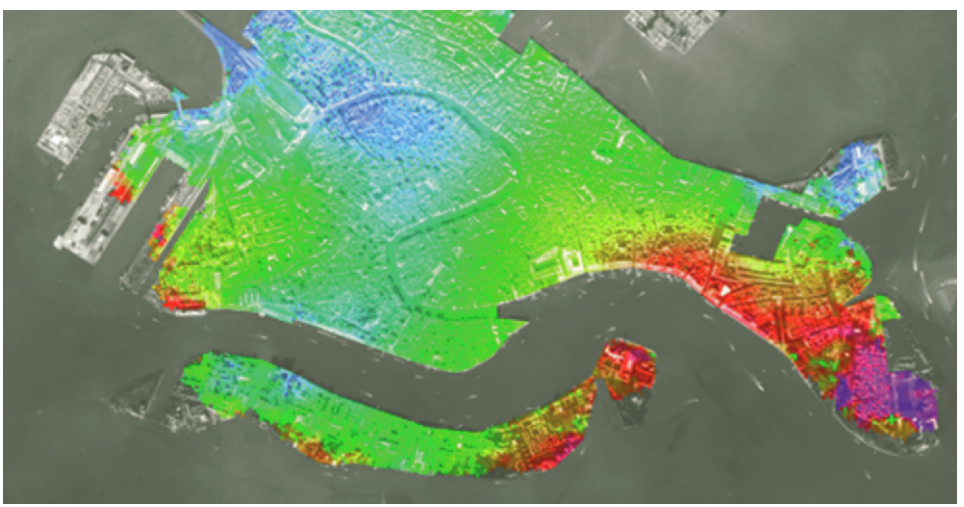


Figure 10.21 This SAR interferometric map of Venice shows up to 2 mm per year (purple) of subsidence over the period 1992–1996. (*Credit*: GAMMA–ESA.)

phenomenon, interpreted as being due to the combined effect of the heating of the magma chamber located between 8 and 16 km deep, by a hot-spot 600 km below, and the internal hydrothermal pressure. Although the phenomenon is not necessarily synonymous of an imminent eruption it calls attention to a serious risk. Other hazards either of a natural or of an anthropogenic origin can be studied with millimeter accuracy over areas at the square kilometer scale, such as the subsidence of cities like Venice (Figure 10.21) or New Orleans. The technique is also applied for monitoring humidity and vegetation cover.

10.3.5 Optical imaging

Imaging with radars offers a lot of advantages but cannot replace optical imaging which provides both higher resolution pictures – as a consequence of using much shorter wavelengths – and observations in many different colors, which help to characterize the nature of soil surface or agricultural features. The angular resolution of imaging instruments has constantly increased since the early ages of Earth observation from space and is now reaching the centimeter range. Contrary to SAR imaging, optical imaging is unfortunately constrained by the presence of clouds and to day-time observations. Imaging instruments fulfill several purposes [25]; they provide observations of the oceans, the land, the cryosphere and also the clouds and the atmosphere. In combination, they allow the monitoring of global climatological and environmental evolution. In addition, they allow the measurement of biological and physical variables of the ocean; in particular the amount of phytoplankton and land cover (Figures 10.8 and 10.22). They are now offering an unprecedented capacity in the prevention and forecasting of environmentally critical situations. Space data are cross-compared with ground-based data and are introduced in computer models that are able to analyze and identify the changes and their evolution. Very promising, for example, is the use of meteorological satellite images to follow and forecast the spread of diseases transported by pets and mosquitoes as they sense climatic variations. Africa and India are particularly vulnerable and are considering the use of this technique, promoted by the French space agency CNES, to track the progression of cholera, malaria and of the hemorrhagic dengue fever. A particularly important application of imaging data in conjunction with ground-based information is the mitigation of famines. A Famine Early Warning System Network was set up in the Sahel regions of Africa and is now in operation in other arid zones of the developing world.

Space remote-sensing optical imaging also allows for the monitoring of volcanoes and earthquakes, as well as anthropogenic hazards. Their power is considerably increased by combining their data with those of other imaging instruments such as the SARs which proved to be ideal tools for monitoring the thermal status of the Earth's surface during the night and in getting an early warning of any imminent eruptive activity [26]. At several occasions these instruments have also been able to follow the progression and regression of forest and land fires. The proper monitoring of fires is particularly crucial for the preservation of the vegetation and the study of the carbon cycle. ENVISAT [27]



Figure 10.22 The Japanese ALOS satellite captured this image over Cardiff, Wales, on 15 June 2006, with its Advanced Visible and Near Infrared Radiometer type 2 (AVNIR 2), which is designed to chart land cover and vegetation in visible and near infrared spectral bands, with a resolution of 10 meters. (*Source*: ESA; *Credit*: JAXA.)

has provided remarkable pictures of the fires that occurred in Borneo in 2002, Portugal in 2005, and Greece and California in 2007 (Figure 10.23). The combination of several instruments has led to important advances: the Advanced Synthetic Aperture Radar (ASAR) can pierce through smoke clouds to provide high-resolution fire impact assessment while the Medium Resolution Imaging Spectrometer (MERIS) provides large-scale fire scar mapping. The sources of the flames can then be spotted by the Advanced Along Track Scanning Radiometer (AATSR).

Imaging through selected band passes allows filtering spectral signatures of different agricultural products, providing detailed information on agriculture and soil. Observations of changes in land cover, and crops monitoring over large areas, as well as productivity forecasting, are ideally suited to space-based multi-



Figure 10.23 This ENVISAT image of wildfires in southern California was acquired on 22 October 2007. It also captures fierce easterly winds blowing dust out from the desert. These fires forced the evacuation of a quarter of a million people. (*Credit*: ESA.)

color imaging and photometry. A very peculiar use of this technique has been introduced for police and security reasons. Since 1989, using the data from SPOT (which are provided commercially), the US Office of the Narcotics Control Board, in collaboration with the Asian Institute of Technology, has implemented an annual control of poppy cultures in Thailand, which has succeeded in drastic reductions of the cultivated areas inside the so-called Golden Triangle (Figure 10.24). The European Union is using space imaging data to ensure that farmers are properly implementing the regulations dictated by the Common Agricultural Policy. All aid applications supplied by each individual farmer are checked for eligibility of land use and area declared, and control ensures that aid for any

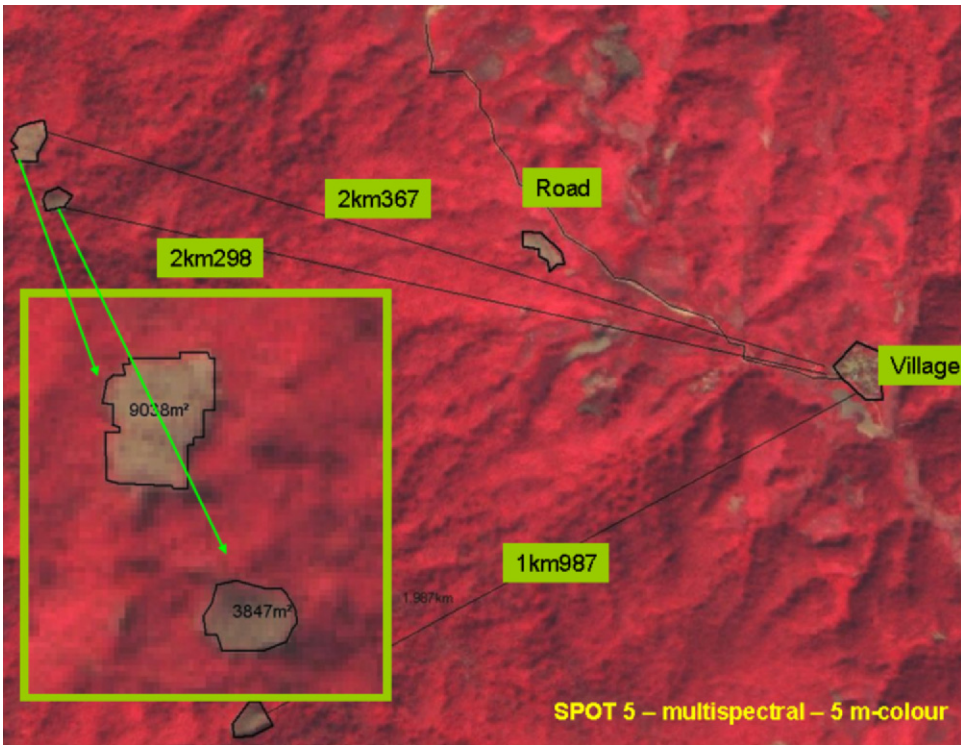


Figure 10.24 Opium poppy parcels delimitation in South-East Asia. (*Credit*: CNES distribution-Spotimage.)

piece of land is claimed only once, and only by one farmer, helping the Union to identify anomalies and frauds. Hyperspectral imagery is a powerful development of that technique which allows the selection of more than 100 spectral bands that can be scanned rapidly and provide improved information on a large number of different types of cultures. Its use, however, certainly represents a major challenge for the analysis and interpretation of large volumes of data.

10.3.6 Remote-sensing spectroscopy

Spectroscopy of planetary atmospheres and surfaces represents one of the most powerful tools for analyzing the chemical state of a planet and its evolution. In the case of the Earth, it is one of the main techniques for the monitoring of global warming and atmospheric pollution over industrial regions and biomass-burning areas. The principle of these measurements is to use the properties of solids or of gases to selectively reflect, scatter or absorb light of different wavelengths, allowing their chemical composition to be inferred. A comprehensive description of remote-sensing techniques is given in [27] and [28].

Obviously, atmospheric observations are a central part of any Earth observation mission. Operating in many different spectral bands, they provide

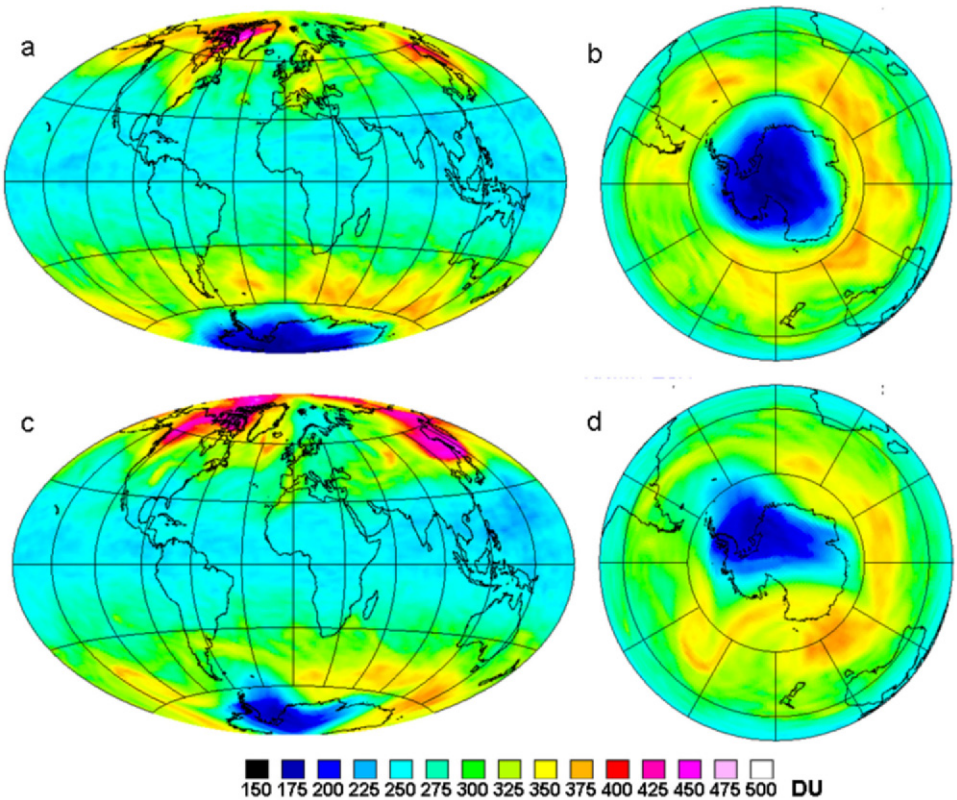


Figure 10.25 Panel (a): Total global ozone column for 16 November 2007 measured by SCIAMACHY. Panel (b): Same for Antarctica. Panels (c) and (d): 8-days forecasts for 25 November 2007. The color codes correspond to different values of Dobson Unit (DU) [30]. (*Credit*: KNMI/ESA.)

global data on a large variety of atmospheric-state parameters, primarily abundance profiles of atmospheric gases (particularly ozone) over Antarctica and the Artic. Atmospheric sounders like the Global Ozone Monitoring Experiment on ERS 2 (GOME) and the Global Ozone Measurement by Occultation of Stars (GOMOS) instrument on board ENVISAT have contributed major advances in the monitoring of ozone depletion. While GOME uses the Sun as a light source, GOMOS uses the light of target stars in the ultraviolet, visible and near infrared ranges as they set closer to the Earth's horizon seen from the satellite [29]. Stars – which are point-like sources of light as opposed to the Sun, which has an angular apparent diameter of half a degree – provide a much higher resolution in the atmosphere. GOMOS allows us to derive the total ozone content in the range 15 to 80 km, with a vertical resolution better than 1.7 km. These maps are used to forecast the evolution of the total ozone content (Figure

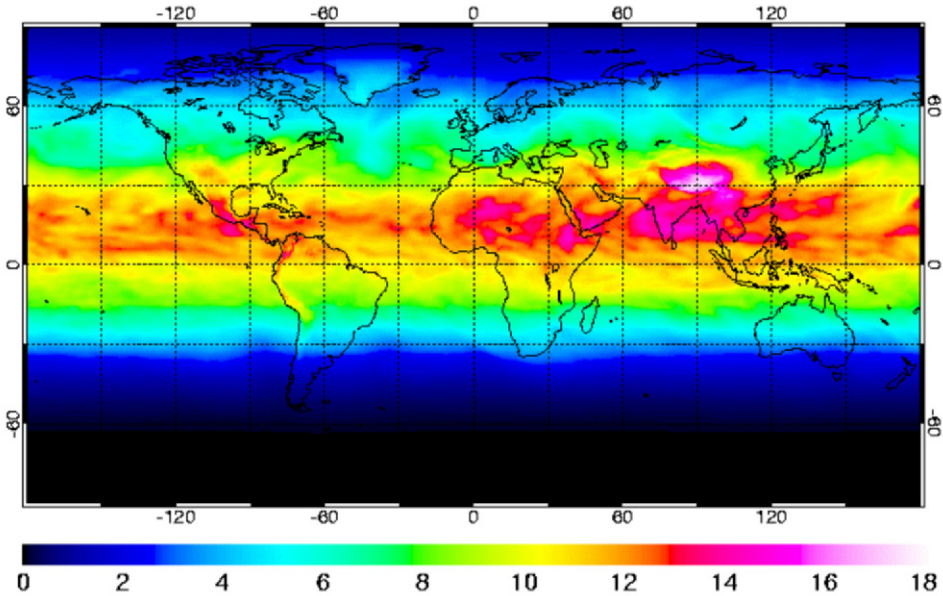


Fig.10.26 Clear sky erythemal UV index for 1 June 2006. Such forecasts are provided regularly by ESA based on SHIAMACHY measurements. (*Credit*: ESA.)

10.25) [30] and to establish regular maps of the clear sky erythemal ultraviolet index [31] (Figure 10.26). Similar instruments are in operation on board other satellites in the USA, Canada and Japan.

Limb occultation and emission measurements combined with nadir observations form the principle of the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) on board ENVISAT [27]. This passive instrument uses the ultraviolet, visible and near-infrared light emitted by the Sun and the Moon that is either transmitted, reflected or scattered by the atmosphere to infer the chemical and physical state of the troposphere, the stratosphere and the mesosphere. The technique is very well suited for the study of aerosols and clouds and of chemical atmospheric compounds. Of most interest are the CFCs, the greenhouse gases and pollution active species such as NO_2, CO, CH_4. NO_y compounds are particularly noxious and deserve special attention. Atmospheric NO_2, for example, is formed from NO, itself the result of the combustion of fossil fuels in cars, trucks, ships, etc. NO_2 is transformed into nitric acid which falls back on the ground and on vegetation, contributing together with other factors to increasing rain and soil acidity. While NO_2 emissions are directly related to the level of industrial activity, NO and NO_2 concentrations have been found to be higher in cities during high business hours. According to the World Health Organization, NO_2 not only induces respiratory difficulties, especially for children, but also contributes to the increasing corrosion of various materials and compounds. Figure 10.27 presents the mean global NO_2 pollution column

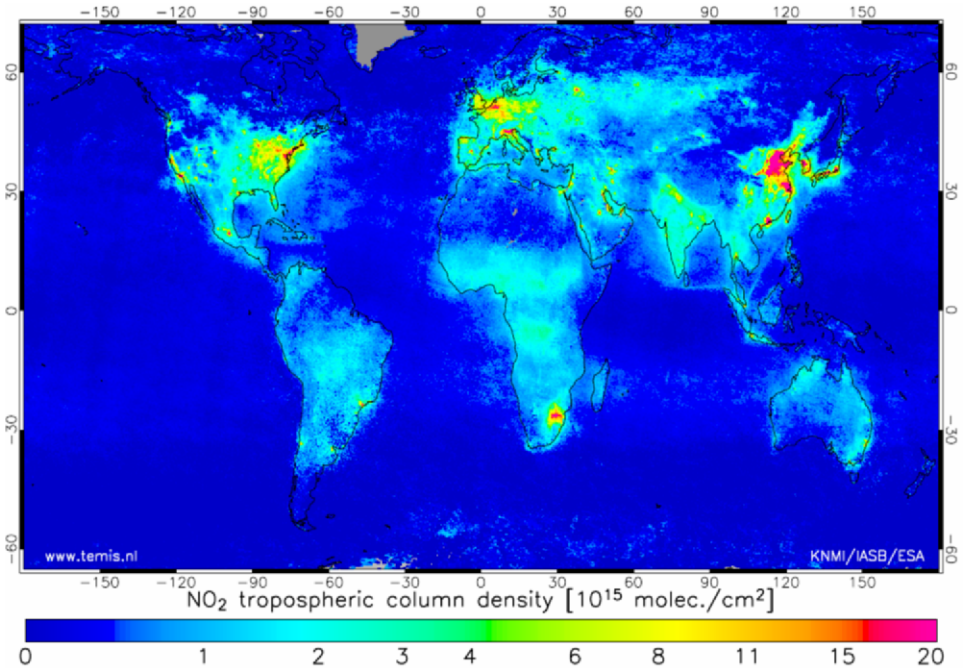


Figure 10.27 Mean tropospheric vertical column density of NO_2 for 2004, as measured by SCIAMACHY in units of 10^{15} molecules/cm^2. Noticeable are the high concentrations in northern Italy and the Benelux as well as in northern America, southern Africa and Asia, especially China. Noteworthy also are the tracks of ships south of India and through the Suez Canal. (*Courtesy*: ESA and University of Heidelberg.)

density map for 2004, measured by SCIAMACHY. Figure 10.28 presents a similar map for methane (2003). SCIAMACHY can also measure the global distribution of CO with uniform sensitivity from the upper atmosphere down to the Earth surface where the CO sources such as industrial activity, fossil fuel burning and forest fires are located.

NASA's Orbiting Carbon Observatory measures visible and infrared sunlight that has traveled through the atmosphere twice: on its way down to the ground and as it is reflected back by the Earth surface. The CO_2 bands are used to measure the absorption and deduce the quantity of CO_2 molecules present in the atmosphere with a precision of nearly one part in 400. A shortcoming is the presence of clouds and the impossibility to perform the measurements at night. The Japanese GOSAT mission also measures CO_2, methane, water and ozone following the same principle using longer wavelengths that allow measurements even in the absence of sunlight, using the infrared emission of the ground avoiding night-time or polar winter data gaps [32].

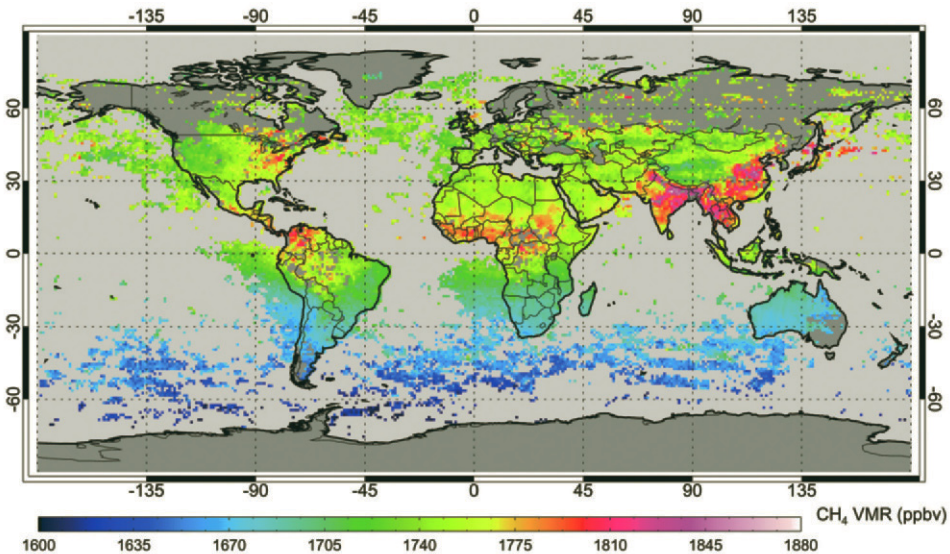


Figure 10.28 Map of the global vertical column density of atmospheric methane derived from data acquired between August and September 2003 by SCIAMACHY. The scale is in parts per billion per volume. The high concentration over the Indian subcontinent and China is due to rice cultivation and agriculture. (*Credit*: ESA and University of Heidelberg.)

10.3.7 Radiometry

The Earth's daily weather and climate is controlled by the balance between the amount of sunlight received by the Earth's surface and the atmosphere, and the amount of energy emitted by the Earth into space. The budget of incoming and outgoing energy is called the radiation budget and is schematized on Figure 10.29. Radiometers measure the amount of energy of all these components. The energy received from the Sun is mostly contained in the visible, while the energy emitted by the surface of the Earth and the clouds is maximal in the infrared and long wavelength radiation. Some of the shortwave radiation from the Sun is reflected back into space by water vapor, ozone, clouds and aerosols. One of the most intriguing questions for climate modeling is how clouds affect the climate and vice versa. Understanding these effects requires a detailed knowledge of how clouds absorb and reflect incoming shortwave energy from the Sun and outgoing long-wavelength radiation emitted by the Earth. An international effort involving France and the USA has materialized with the development of the Cloud–Aerosol Lidar and Infrared Pathfinder Satellite (CALIPSO) whose objectives are exactly to understand the respective roles played by these two atmospheric components in the radiation budget [33, 34].

Radiometry of the Earth's surface is constrained to 'windows' where the atmosphere is nearly transparent (Figure 10.30) [35]. With infrared imaging instruments, radiometry provides temperature measurements of the Earth and

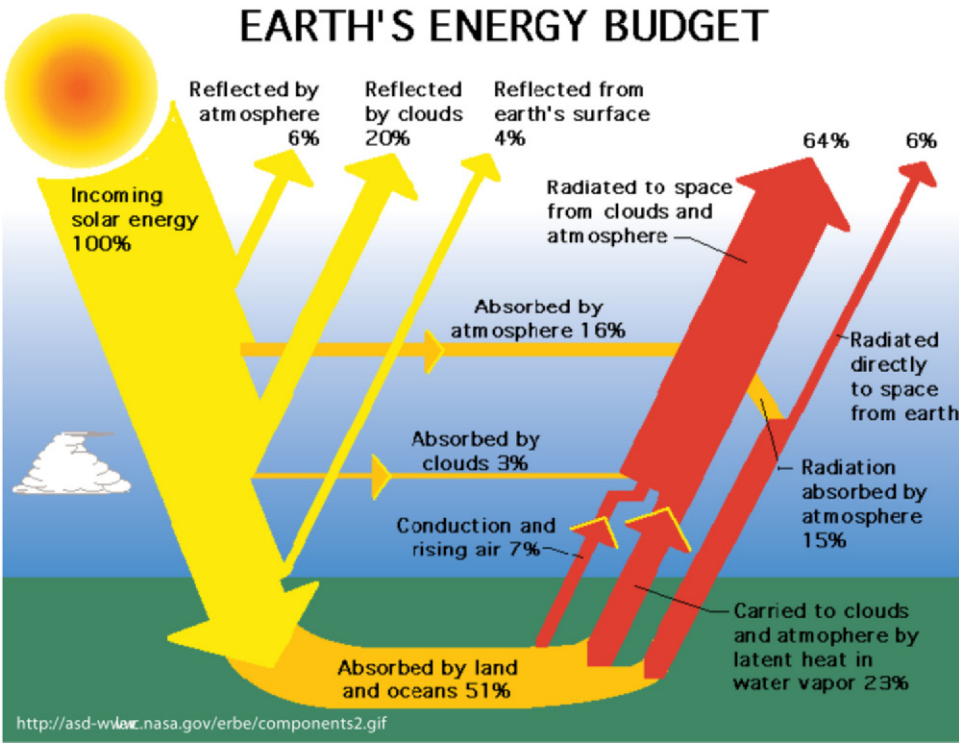


Figure 10.29 The Earth's energy budget balances the energy received from the Sun in the form of infrared, visible and ultraviolet light with that which is absorbed by the land, the oceans and the atmosphere and re-emitted into space by the Earth.

sea surface, reaching a precision of a few tenths of a degree (0.3° with the AATSR on ENVISAT). Such measurements are essential for the monitoring of the greenhouse effect and of natural climatic fluctuations of the global or local temperature (Figure 10.31). They are available continuously owing to the international fleet of Earth observation and meteorological satellites.

A very specific use of radiometry concerns the heating of the Earth by volcanic activity. NASA's satellite MODIS (Moderate Resolution Imaging Spectro-radiometer) measured the heat emitted by the world's 45 most active volcanoes, which throw out about 5×10^{16} joules per year – enough to power New York city for a few months. When Mount St Helens erupted on 18 May 1980, it released more than 10^{18} joules of heat at once – about 20 times the total yearly heat flow from all the volcanoes studied in 2001. While these numbers are relatively small in terms of the Earth's overall heat generation, they contribute to our understanding of the planet's heat flow and are potentially useful for forecasting volcanic activity.

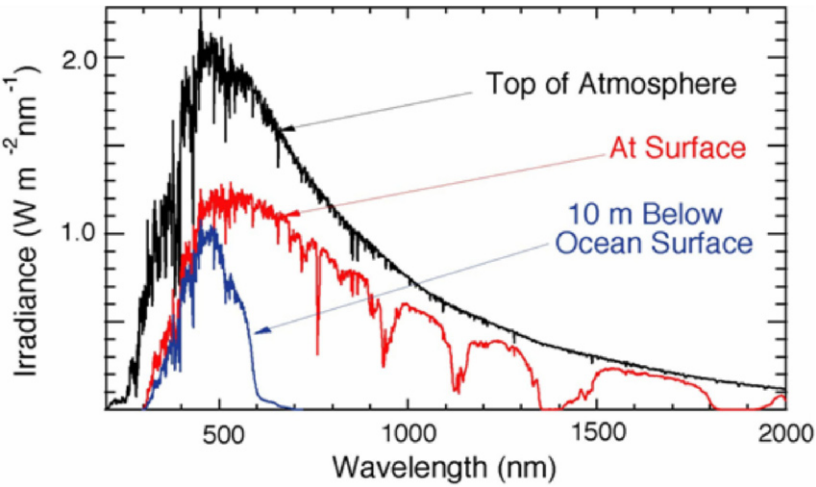


Figure 10.30 The solar spectral irradiance at the top of the atmosphere is compared with the irradiance at the Earth's surface (red) and 10 meters deep in the ocean (blue). Irradiance at the surface shows strong atmospheric absorption bands in the infrared primarily due to water vapor and the complete extinction at wavelengths below 300 nm due to ozone. The strong infrared absorption of water is apparent in the irradiance spectrum penetrating into the ocean [35]. (*Courtesy*: G. Rottman.)

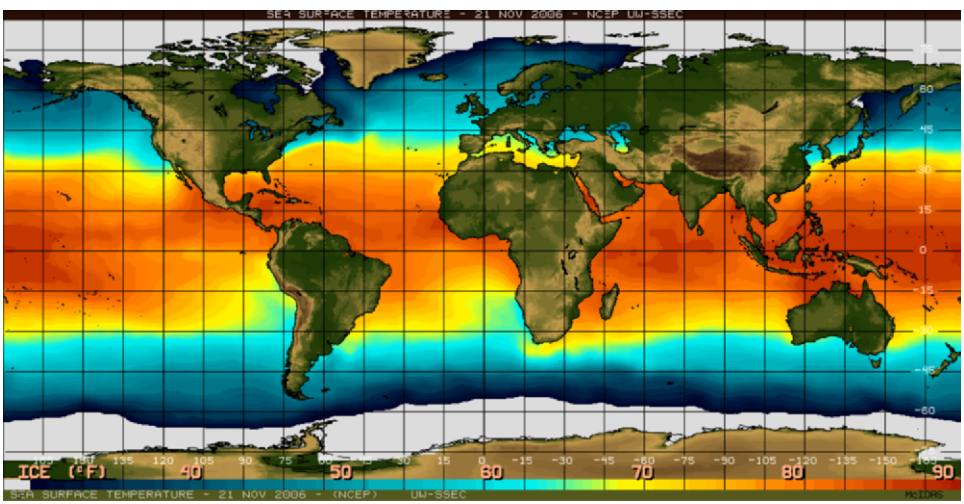


Figure 10.31 Global map of the sea surface temperature of 21 November 2006, ranging from 0°C to 32°C. These maps are established on a daily basis from US NOAA satellite data. (*Credit*: Space Science and Engineering Center, University of Wisconsin, Madison.)

10.3.8 Monitoring astronomical and solar influences

As discussed in Chapter 3, our planet is influenced by the astronomical environment of the whole Solar System. Astronomical satellites are offering the means to quantify these cosmic hazards. In the case of asteroids, satellites are mandatory if mitigation measures are envisaged. Gamma-ray bursts have been discovered by military satellites and it is likely that, in the future, high energy X-ray and gamma-ray astronomical satellites will continue to monitor such events. However, the largest influence is due to the Sun, which means that we must analyze this influence in more depth.

Chapter 5 discussed the relative importance of solar forcing on the Earth climate to conclude that if it is indeed real, it is not the most determining factor. On the contrary, the Sun exerts a strong influence on the upper atmosphere, the troposphere, the stratosphere and the thermosphere above, due to atmospheric absorption of solar ultraviolet light (Figure 10.30). The intensity of that light is strongly dependent on the 11-year solar activity cycle which modulates the number of magnetic sunspots on the solar disk as well as the morphology of the external layers above the disk: the chromosphere and the corona (Figure 10.32)

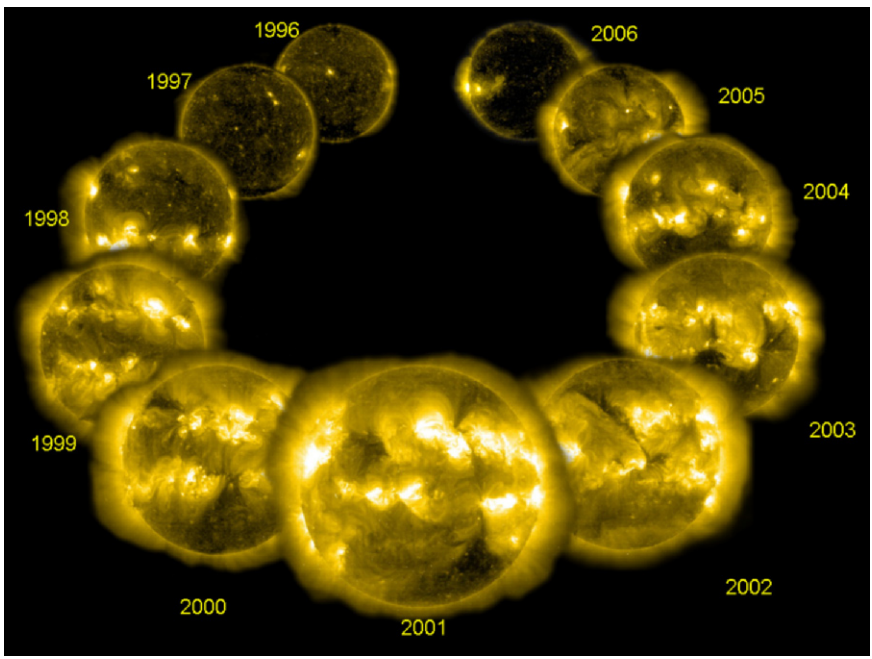


Figure 10.32 The morphology of the Sun's external layers is strongly influenced by the solar cycle as shown on this series of pictures obtained with the EIT instrument [36] on board the SOHO satellite over a period of a complete solar cycle. Solar maximum was reached in 2001 and minimum in 1996 and 2007. The wavelength of 284 nm is in the extreme ultraviolet, and the corresponding radiation is emitted by the corona at a temperature of 2 million degrees. (*Credit*: SOHO/EIT (ESA and NASA).)

[36]. Because of atmospheric absorption, ultraviolet solar irradiance can only be measured from space. More than 30 years of such measurements have shown that solar variability is increasing in the far ultraviolet, as the chromosphere and the corona are the predominant contributors to ultraviolet and X-ray emission, and the shorter the wavelength, the stronger is the variability (Figure 10.33).

At wavelengths shorter than 242.2 nm, molecular oxygen, O_2, is photodissociated, liberating two oxygen atoms. This process is the key step in the formation of the ozone stratospheric layer as the oxygen atoms thus created combine with O_2 to create ozone, O_3. Ozone is itself destroyed by solar ultraviolet light at wavelengths shorter than 310 nm in the so-called Hartley band. At these wavelengths solar UV radiation is less sensitive to solar activity (Figure 10.33) [37, 38], hence ozone production is more strongly modulated by solar activity than its destruction, and the net production is higher during periods of high solar activity. As enhanced concentrations of stratospheric ozone result in greater absorption at wavelengths shorter than 310 nm and longer than 500 nm, less solar radiation reaches the troposphere at solar maximum resulting in a slight cooling of these layers. However, these effects are strongly non-linear, and the estimates of their impacts on the terrestrial climate system vary widely [39].

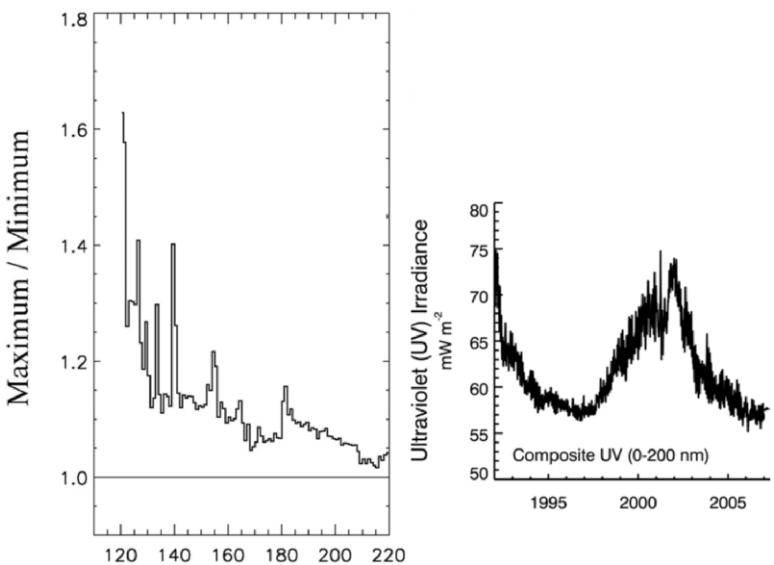


Figure 10.33 *Left panel*: Solar cycle variations of the UV Solar Spectral Irradiance observed by the UARS mission (NASA), from solar maximum in 1992 to solar minimum in 1996. The wavelength-by-wavelength ratios of the maximum to minimum are plotted as a function of wavelength measured in nanometers (nm) and values range from less than 1% in the near ultraviolet to almost a factor of 2 at wavelengths shorter than 100 nm [37]. *Right panel*: The ultraviolet irradiance integrated over 0–200 nm is plotted as a function of time between 1992 and 2007 evidencing the strong modulation by the solar cycle (adapted from reference [38].)

At solar maximum, the upper layers of the Earth's atmosphere up to some 600 km or even higher are heated by the excess ultraviolet radiation, resulting in changes of the density and an increased friction on satellites that experience a faster loss of altitude in Low Earth Orbit. The extreme ultraviolet and X-ray radiation is also responsible for the formation and heating of the ionosphere through the ionization of oxygen and nitrogen atoms (see Box 10.4). The ionosphere (Figure 10.4) plays a major role in the transmission of radio waves and of electromagnetic signals in general. This effect is particularly troublesome for satellite communications. When the ionosphere between satellites and the user becomes turbulent and irregular, the signal may 'scintillate' and is more difficult to track. For example, the Total Electron Content along the path of a GPS signal can introduce a positioning error of up to 100 meters. This is one of the most significant effects of the so-called 'space weather' due to the planned reliance on the GPS in the future and presently used by several hundred million people around the world.

The term 'space weather' is used to describe the influence of the Sun on interplanetary space and on the Earth. Regarding the Earth, it is a consequence not only of the behavior of the Sun but also of the Earth's magnetic field. It is greatly influenced by the speed and the density of the solar wind and by the interplanetary magnetic field carried by the solar wind. A variety of physical phenomena are associated with space weather, including geomagnetic storms and substorms, energization of the Van Allen radiation belts, ionospheric disturbances and scintillation, aurora and geomagnetically induced currents at Earth's surface. At solar maximum the Sun also emits a larger number of flares, coronal mass ejections (CME) and highly energetic particles (mostly nuclei of hydrogen atoms). These so-called solar–proton events do have an effect on the chemical composition of the upper atmosphere, especially on ozone which is depleted and on NO_2 which increases according to space observations made, in particular, by GOMOS [40, 41]. CME and their associated shock waves are important drivers of space weather as they can compress the magnetosphere and trigger geomagnetic storms. Solar energetic particles accelerated by these CME can damage electronics on board spacecraft and threaten the life of astronauts. Therefore, it is important to forecast their occurrence, if at all possible. As their propagation velocities are relatively slow (<500 km/s [42]) compared to the velocity of light, any satellite properly equipped and located between the Sun and the Earth would provide an early warning system. SOHO, placed on a halo orbit around L_1, has indeed played that role.

An indirect connection between solar activity and the Earth's climate has been proposed to act via cosmic rays. Cosmic-ray fluxes are the main cause of the ionization of the atmosphere below about 60 km. Thus, cosmic rays could possibly result in the electrification of aerosols, increasing the probability that they become condensation nuclei. A connection between the rate of formation of low-altitude clouds and the flux of cosmic rays has indeed been invoked via the interaction with aerosols [43]. Inversely to solar ultraviolet radiation, cosmic rays penetrate less easily in the atmosphere at solar maximum because the

Box 10.4 The ionosphere

The ionosphere (Figure 10.4) is the uppermost part of the atmosphere at altitudes of 70–400 km and above, which is ionized by solar radiation and becomes conductive. It plays an important part in atmospheric electricity and influences radio propagation to distant places on the Earth. It is made of different layers corresponding to ionization by ultraviolet and X-ray solar radiation of different wavelengths.

The D layer, the innermost layer, extends 70 to 90 km above the surface of the Earth, and is due essentially to the ionization of nitric oxide (NO) by the light of the solar hydrogen resonance line – called Lyman alpha – at the wavelength of 121.6 nm. The D layer is mainly responsible for absorption of high-frequency (HF) radio waves, particularly at 10 MHz and below, with progressively smaller absorption as the frequency gets higher. The absorption is small at night and greatest about midday. The layer reduces greatly after sunset, but remains due to the ionization effect of galactic cosmic rays.

The E layer extends 90 to 120 km above the surface of the Earth and is due to the ionization of molecular oxygen (O_2) by soft X-rays and EUV radiation (1–10 nm). This layer reflects radio waves of frequencies less than about 10 MHz. At night the E layer begins to disappear because the primary source of ionization is no longer present. This results in an increase in the height where the ionization is maximal because oxygen atoms recombine faster in the lower layers. The E_s layer or sporadic E layer is characterized by small clouds of intense ionization, which can support radio wave reflections from 25 to 225 MHz. Sporadic E events may last for just a few minutes to several hours and open up propagation paths that are generally unreachable.

The F layer extends 120 km to 400 km above the surface of the Earth, and is due to the ionization of atomic oxygen (O) by EUV radiation (10–100 nm). The F layer is the most important part of the ionosphere in terms of HF communications. Acting like a mirror, it is mostly responsible for the propagation of radio waves around the Earth. Most long-distance HF radio communications (between 3 and 30 MHz) propagate thanks to the presence of the F layer rather than being lost in space.

For all these layers, and more predominantly for the E and F layers, there is a clear solar cycle effect, the average electron and ion densities being higher during solar maximum.

stronger magnetic fields act like a shield [44] (see also Chapter 5). That could explain a negative correlation between solar activity and cloud formation, which has still to be proven, however.

Given the importance of solar variability on several elements of the Earth's system – in particular the upper layers of its atmosphere – it is of great importance that the most crucial parameters such as the total and spectral solar

irradiance, the sunspot cycle, the geomagnetic activity and the solar particle and cosmic ray fluxes, be continuously monitored in view of establishing a database that will help to separate solar forcing from the anthropogenic perturbations, and assessing their evolution and the efficiency of any corrective measure. Space weather monitoring and forecasting require a substantial number of missions operated in a coordinated way. Table 10.1 summarizes the essential needs for a complete space weather system. An ideal set of monitors would include several satellites watching the Sun, measuring its radiation and all manifestations of its activity located between the Sun and the Earth. They would track disturbances from the Sun to the Earth and provide early warning of the arrival of CMEs and high-energy solar proton events. Close to Earth, constellations of small satellites placed in key regions of geospace would monitor the magnetosphere and geomagnetic perturbations, while satellites like Swarm would monitor the Earth's magnetic field.

Several satellites still in operation for relatively long periods of time fulfill at least part of the expected duties but not necessarily all of them, and not in a continuous mode of operations: SOHO (ESA–NASA), Ulysses (ESA–NASA), ACE (NASA), STEREO (NASA), SORCE (NASA), TIMED (NASA), HINODE (Japan), CLUSTER (ESA–NASA), 'Double Star' (China), IMAGE (NASA), MAGSAT (Canada), 'Oersted' (Denmark), DEMETER (CNES). The National Space Weather Program Council of the United States has prepared one of the most complete plans for a Space Weather Program which describes the needs and the

Table 10.1 An ideal set of missions for a space weather observation system, ranging from solar observatories (visible, ultraviolet, X-rays), Out-of-Ecliptic or solar-polar missions, satellites at Lagrange point L_1, in the magnetosphere or in GEO and LEO.

Sun	Interior: Helioseismology	Polar missions
	Photosphere: Visible radiometry	Visible-total light radiometry
	Magnetic fields: sunspots	Magnetograms, Imagery
	Chromosphere: loops, flares, activity	UV and EUV spectral-imagery
	Corona: loops, flares, CMEs, activity	EUV, X-rays, radio
Solar wind	CMEs	Satellites at L_1
	Solar Particles	Out of Ecliptic missions
	Proton events	Particle detectors
	Cosmic rays	
Magnetosphere	Magnetic fields	3-D measurements
	Geomagnetic disturbances	Multi-point missions
	Radiation belts	
Ionosphere	Total Electron Content	Imaging instruments
	Auroras	Ground Based Sounders
	Electric fields	GPS, Geostationary S/C
	Disturbances/Scintillation	Spectrometry
	Neutral atmosphere	

requirements for such a program [45]. NASA's Living With a Star (LWS) Program seeks to advance understanding of solar variability and its effect on Earth. The program consists of an observational portion, based on dedicated missions, and a supporting theory and modeling program. Other organizations such as ESA are also discussing similar plans and have expressed an interest in an internationalization of such elements in order to get a system close to an ideal configuration of missions. The international commitments, however, are not binding and the system may not contain all the most desirable instruments. This is left in the hands of goodwill or individual interests.

The need for continuity of the measurements is another important aspect of the requirements. It is in no way guaranteed that there will always be a satellite in orbit with the proper instrumentation to provide the required uninterrupted data sets with the proper level of accuracy. For example, in the period 1980–1997, no measurements were available in the XUV/EUV range, making it more difficult to correlate solar variability and its possible effects on Earth or to reconcile discontinuous measurements [37].

10.4 Conclusion

From what precedes, it is clear that artificial satellites are becoming every day more indispensable for assessing the evolution of the Earth and of the environment. Their increasing observing power, their accuracy and the precision of their measurements, their role in analyzing and mitigating natural and anthropogenic hazards has been proven and is in constant progress. Unfortunately, the critical infrastructure to make space-based data available for a proper management of the planet and for the security of its inhabitants is not complete, even though the elements that already exist are showing every day their potential power for helping the scientists to understand the complexity of the Earth system and possibly to forecast its evolution.

More than 170 Earth observation satellites are in operation under the responsibility of the main space agencies of the world, providing essential information to scientists and politicians in their respective countries and on a broad international basis. This is essential, given the global character of the problems they are able to address. But are they sufficient in type or in number? A substantial number of scientific and international organizations develop or coordinate some of the necessary space and ground-based tools, but which of them is globally responsible for ensuring that these tools are available when required, and are maintained to ensure their indispensable continuity to properly assess the evolution of the Earth on the medium and long-term scales? There are signs that even the incomplete current capability is in jeopardy. This is the subject of the following chapter.

10.5 Notes and references

[1] *The Changing Earth*, 2006, ESA Special Publication SP-1304, p. 85. See also the ESA website: http://www.esa.int/esaEO/

[2] Lockwood, M., 2006, 'What do cosmogenic isotopes tell us about past solar forcing of climate?' in Calisesi, Y. and Bonnet, R.M., (eds), *Solar Variability and Planetary Climates* (ISSI Space Science Series **23**), and *Space Science Reviews* **125** (1–4), 95–109.

[3] The sinking or rising of the ground due to modifications induced underground by, for example, pumping of oil or water.

[4] Murray, T., 2006, 'Greenland's ices on the scales', *Nature* **443**, 277–278.

[5] Wingham, D., 2005, 'Cryosat: A mission to the ice fields of Earth', *ESA Bulletin* **122**, 11–17.

[6] Cazenave, A., 2003, *Observations depuis l'espace de la terre solide, de l'Océan et des eaux continentales,* Académie Nationale de l'Air et de l'Espace, Les apports de l'espace dans le progrès de la gestion humaniste de la Planète, Toulouse 27 Nov. 2003, Proceedings.

[7] The scale height is the vertical distance over which the density decreases by a factor of 2.718, called Euler's number.

[8] *Environmental effects of ozone depletion and its interaction with climate change,* United Nations Environment Program, 2006 Assessment, and *Twenty questions and answers about the ozone layer,* a Panel Review Meeting for the 2002 ozone assessment led by W. Fahey, Les Diablerets, Switzerland, 24–28 June 2002.

[9] Meinrat, O.-A. et al., 2005, 'Strong present-day aerosol cooling implies a hot future', *Nature* **435**, 1187–1190.

[10] Liebig, V., 2005, *CEOS Earth Observation Handbook*, ESA, www.eohandbook.com, p. 212.

[11] A. Gore, former US Vice-President who shared the Peace Nobel Prize 2007 with the IPCC, had suggested in the late 1990s to develop a solar and Earth satellite located in L_1 for monitoring the Earth radiative budget and the climate. The project never was implemented partly because of the changes of US President in January 1981 and because no organization was assigned the responsibility of the development.

[12] The unit is named after Galileo Galilei (1564–1642) who proved that all objects at the Earth's surface experience the same gravitational acceleration.

[13] Horwath, M. and Dietrich, R., 2006, 'Errors of regional mass variations inferred from GRACE monthly solutions', *Geophysical Research Letters* **33**, Issue 7, L07502.

[14] Chambers, Don P., 2006, 'Evaluation of new GRACE time-variable gravity data over the ocean', *Geophysical Research Letters* **33**, Issue 17, L17603.

[15] ESA Brochure BR-209, June 2006, p. 17.

[16] Barlier, F. and Lefebvre, M., 2001, 'A new look at Planet Earth: satellite geodesy and geosciences', in Bleeker, J.A.M. et al. (eds), *The Century of Space Science,* Kluwer Academic Publishers, Vol**. II**, 1623–1651.

[17] Benveniste, J. et al., 2001, 'The radar altimetry mission: RA-2, MWR, DORIS and LRR', *ESA Bulletin* **106**, 67–76.
[18] Ward, S., 2002, 'Slip-sliding away', *Nature* **415**, 973–974.
[19] Limb sounding measures the excess delay of radio waves as they are affected by refraction when they pass through the Earth's atmosphere. The precise degree of bending (and hence the precise excess delay) is highly dependent on atmospheric pressure, temperature, and moisture content which can then be estimated from precise observations of the changing signal delay as the GPS satellite rises or sets.
[20] 'Spaceborne radar applications in geology. An introduction to imaging radar and application examples of ERS SAR', in K. Fletcher, (ed.) *Geology and Geomorphology*, 2005, ESA TM-17.
[21] Pritchard, M.E., 2006, 'InSAR, a tool for measuring Earth's surface deformation', *Physics Today* **59** (7), 68–69.
[22] Fialko, Y. et al., 2005, **'**Three-dimensional deformation caused by the Bam, Iran, earthquake and the origin of shallow slip deficit', *Nature* **435**, 295–299.
[23] Massonnet, D. et al., 1995, 'Deflation of Mount Etna monitored by spaceborne radar interferometry', *Nature* **375,** 567–570.
[24] Massonnet, D. and Feigh, K.L., 1998, 'Radar interferometry and its application to changes in the Earth's surface', *Reviews of Geophysics* **36** (4 November), 441–500.
[25] Huot,J.-P. et al., 2001, 'The optical imaging instruments and their applications: AATSR and MERIS', *ESA Bulletin* **106**, 56–66.
[26] Zink, M. et al., 2001, 'The radar imaging instrument and it applications', *ESA Bulletin* **106**, 46–55.
[27] Nett, H. et al., 2001, 'The atmospheric instruments and their applications: GOMOS, MIPAS and SCIAMACHY', *ESA Bulletin* **106**, 77–87.
[28] Rees, W., 2001, *Physical Principles of Remote Sensing,* Cambridge University Press, p. 335.
[29] The measurement principle consists in pointing the satellite towards the Earth's horizon as seen from orbit, allowing the instruments to observe the dimming of the light from the Sun or from a star as they cross thicker and thicker layers of the atmosphere. Atmospheric transmission is computed at all wavelengths as the ratio between the absorbed spectra as the star sets behind the horizon and the undisturbed spectrum of the target star of the Sun, detected at tangent heights above the atmosphere.
[30] A total value of 500 DU is equivalent to a layer of pure ozone of 0.5 cm at the Earth's surface and a temperature of 0°C; 300 DU correspond to a layer of 0.3 cm.
[31] Sun-burning UV radiation or 'erythemal radiation' is harmful to humans and other life forms. The erythemal UV index is set to zero when darkness occurs such as at high latitudes during local winters. It is the highest in the tropics in summer time where and when the Sun is close to vertical.
[32] Haag, A., 2007, 'The crucial measurement', *Nature* **450**, 785–787.

[33] Ramanathan, V. et al., 2007, 'Warming trends in Asia amplified by brown cloud solar absorption', *Nature* **448**, 575–578.
[34] Pilewski, P., 2007, 'Aerosols heat up', *Nature* **448**, 541–542.
[35] Harder, G. et al., 2005, 'The Spectral Irradiance Monitor: Scientific Requirements, Instrument Design, and Operation Modes', *Solar Physics* **230,** 141–167.
[36] Delaboudinière, J.P. et al, 1995, 'EIT; The Extreme-Ultraviolet Imaging Telescope for the SOHO mission', *Solar Physics* **162,** 291–312; Moses, D. and Clette, F., 1997, 'EIT observations of the Extreme-ultraviolet Sun', *Solar Physics* **175**, 571–599.
[37] Haigh, J.D., 1996, 'The impact of solar variability on climate', *Science* **272**, 981–984; Haigh, J.D., 1999, 'Modeling the impact of solar variability on climate', *Journal of Atmospheric and Solar Terrestrial Physics* **61**, 63–72; Haigh, J.D., 2004, *Climate Lectures*, 34th Saas-Fee Advanced Course of the Swiss Society for Astrophysics and Astronomy, 15–20 March 2004, Davos, Switzerland.
[38] Rottman, G. J. et al., 2004, 'Measurement of solar ultraviolet irradiance', in Pap, J. and Fox, P. (eds), *Solar Variability and its Effect on the Earth's Atmosphere and Climate System*, AGU Monogram **414**, 111–126.
[39] Woods, T.N. and Lean, J., 2007, 'Anticipating the next decade of Sun-Earth system variations', *Eos* **88** (44), 457–458.
[40] Hauchekorne, A. et al., 2006, 'Impact of solar activity on stratospheric ozone and NO2 observed by GOMOS-ENVISAT', in Calisesi, Y., and Bonnet, R.M. (eds), *Solar variability and planetary climates* (ISSI Space Science Series **23**), and *Space Science Reviews* **125** (1–4), 393–402.
[41] Jackman, C. H. et al., 2006, 'Satellite measurements of middle atmospheric impacts by solar proton events in solar cycle 23', in Calisesi, Y. and Bonnet, R.M. (eds), *Solar Variability and Planetary Climates*, (ISSI Space Science Series **23**), and *Space Science Reviews* **125**, (1–4), 381–391.
[42] Hudson, S. et al., 2006, 'Coronal Mass ejections: overview of observations', in Kunow, H. et al. (eds), *Coronal Mass Ejections* (ISSI Book Series **21**), 13–30, and *Space Science Reviews* **123**, (1–3), 2006.
[43] Marsh, N. and Svensmark, H., 2000, 'Cosmic rays clouds and climate', *Space Science Review* **94** (1–2), 215–230.
[44] Scherer, K. et al., 2004, 'Long-Term Modulation of Cosmic Rays in the Heliosphere and its Influence at Earth', *Solar Physics* **224**, 305–316.
[45] National Space Weather Program, *Implementation Plan, 2nd Edition,* Prepared by the Committee for Space Weather for the National Space Weather Program Council, Office of the Federal Coordinator for Meteorology, FCM-P31–2000, Washington, DC, July 2000.

11 Managing the Planet's Future: Setting-Up the Structures

The strongest is never strong enough to be always the master unless he transforms strength into right and obedience into duty.

Jean-Jacques Rousseau

11.1 Introduction

The assumption that humans will survive on Earth for the next 100,000 years is incompatible with our present modes of living. It implicitly requires severe changes in the way the planet is managed on the global scale. The word 'global' is important as it puts into focus the essential need of coordinating all means of management world wide, be they scientific, technical, environmental or political.

The first step is to understand how the Earth system works and how its multitude of components interact. This is a task resting essentially on a rigorous scientific analysis aiming at presenting and formulating the facts in the most understandable and undisputable way to the political world to persuade it to act and implement corrective measures where necessary, as inconvenient as these might be. In this process, the limits of knowledge, of comprehension of phenomena and of uncertainties must be outlined as clearly as possible. This is without doubt the responsibility of the scientific community, and constitutes what we call the 'alert phase'. This alert is directed to the governments, the decision makers and the responsible international organizations. To serve this phase and assist the scientific work, securing also the continuous monitoring of the effects of the corrective measures, the availability of the necessary tools should be guaranteed.

These tools should offer the possibility of continuously observing and permanently monitoring the planet with the support of expanded scientific research programs, in view of taking the necessary political decisions at both regional and global levels. Besides the essential ground-based monitoring systems, satellites offer the most powerful set of technical means because, as discussed in the previous chapter, they are not only able to address global problems but they are also in constant progress, always more accurate and precise. However, the complexity of the Earth system cannot be addressed with

just one or even a few satellites; it requires that a large number of parameters be monitored from different platforms at different locations, different latitudes, and different times, over long periods. It also requires a network of coordinated satellites and instruments. This necessity is just now being perceived, but the tools are not yet systematically organized to respond to the needs whenever identified. In addition, satellite observations must be complemented by research work conducted from the ground, using models and laboratory work.

In a second step, the alert phase should be followed by a political phase during which the scientifically justified measures ought to be implemented. That is probably very theoretical as there is no unique government but a multitude of authorities with different agendas and different interests. Hence the scientific community has also the duty of presenting the political world, and the public world wide, with the consequences and effects if no action is taken. At the end, a new governance of the planet is probably an unavoidable necessity.

11.2 The alert phase: need for a systematic scientific approach

11.2.1 Forecasting the weather: the 'easy' case

Admittedly, forecasting 1,000 centuries looks foolish, as so many factors have to be considered: scientific, natural and anthropogenic hazards, political, economical and societal. In that connection, weather forecasting, even though it has taken centuries to put on a rational and reliable basis, is one of the most simplified cases of forecasting, as only a limited set of factors enter into consideration. That example is offering some hope that more complex cases, such as climate forecasting that we discuss below, or the evolution of resources, biodiversity, and several others, can now be addressed. It is one of the best examples of a scientific approach based on the use of the combined power of satellites, computers and modeling. Weather forecasting over several days requires spatial and temporal fluctuations of the weather to be reliably reproduced with calculations. This rests on models that are very sensitive to the initial conditions: a small error today will lead to much larger ones in a few days and the difficulty increases with the time lapse of the forecast. The advent of space techniques has been significant in improving the accuracy of the data over large geographical areas. Figure 11.1 illustrates the remarkable improvements that have been achieved over the last two decades in the accuracy of weather forecasting over 3, 5, 7 and 10 days.

Weather forecasting using satellites started in different parts of the world through a preliminary scientific investment involving specialists in different fields of physics, and most developed countries, in particular those possessing a space capability, have followed this approach. The first geostationary meteorological satellite was launched by the United States in 1966. Still in the United States, the National Center for Atmospheric Research, NCAR, has formed generations of meteorology scientists and has been a school for many world wide. NCAR and university scientists work together there on research topics in

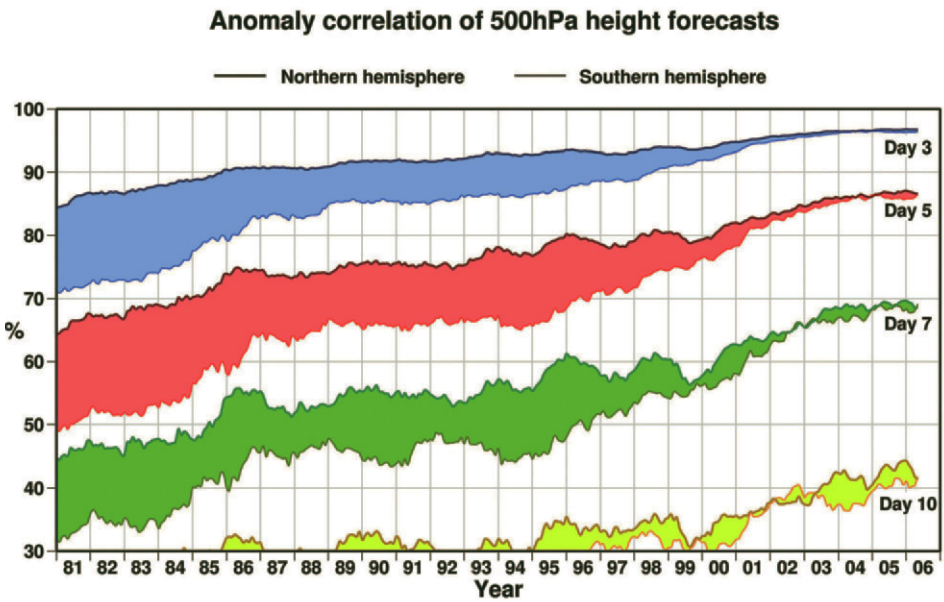


Figure 11.1 Anomaly correlations of 500 Hpa height weather forecasts illustrating the improvement over the last two decades in the reliability of forecasts over 3, 5, 7 and even 10 days. In spite of the decline of part of the *in-situ* observing capacity, skills of the medium-range forecasts have remarkably improved in recent years due to the progress made in the numerical modeling and to the availability of space data: this explains, in particular, the rapid convergence of the northern hemisphere and southern hemisphere forecasting skills that were poorly covered in the past with *in-situ* measurements [1]. (*Credit*: WMO).

atmospheric physics, chemistry, cloud physics and storms, weather hazards and Sun–Earth interactions. In Europe, the first Meteosat spacecraft was initiated in the mid-1960s at the Service d'Aéronomie and the Laboratoire de Météorologie Dynamique. Its leading scientists were formed and trained at NCAR. Taking example from their American colleagues, they realized the need to develop a system of tools based on satellite data, on the use of the most powerful computers and on the development of atmospheric models. That initial 'scientific phase' was followed by the integration in the mid-1970s of the Meteosat program into the European Space Agency, and by the creation of a specific operational organization, Eumetsat, which now provides weather forecasting data to all meteorological offices in Europe. The Japan Meteorological Agency provides forecasts not only for volcanoes and tsunamis but also for short-term alerts on earthquakes.

Today, the data from meteorological satellites are being used operationally by weather services worldwide. Around a third of the planned Earth observation missions can be qualified as having meteorology as a primary objective. As weather knows no national boundaries, international cooperation on the world

scale is an obvious necessity. Forecasting the near and medium term evolution of the weather corresponds to one of the most pressing needs of all nations and people – the rich as well as the poor, those who live in cities and those who live in the countryside, whether in the north or in the south! For that reason, international cooperation is relatively easier to implement in meteorology than in other areas. Since the forecasts are rather short term, they do not require major global political decisions, which is another reason why the development of meteorology has been relatively free of political interference. In that connection, the World Meteorological Organization in Geneva, WMO (see Box 11.1) offers a unique example: it provides the framework for an efficient international cooperation; it has developed mechanisms that enable the provision of forecasts and warning services to all areas of the world, in particular those that suffer from weather-related disasters.; and it is a key player in the successful development of weather and climate forecasting.

Figure 11.1 not only illustrates the progress accomplished and the state of the art in weather forecasting, but also shows the difficulties of improving the time lapse of reliable forecasts. Future progress will come soon from better or more complete observations including, in particular, the measurement of the variation of wind velocities with altitude, as foreseen on ESA's ADM–Aeolus satellite, cloud cover, aerosols, air and sea surface temperature, humidity profiles, etc. (A complete list of all the needs in this domain is found in references [1] and [2].) Improvement will also come from better models. Current ones have a resolution of 30–50 km and by 2025 it is expected to reach the 1-km resolution range owing to better data and computing power. Resolution is a key factor in improving the realism of the precipitation patterns and cyclone forecasts. Future progress will rely more and more heavily on the continuous involvement of scientists in improving the accuracy and reliability of the predictions. Many specialized scientific organizations or institutes around the world are involved in this discipline. This is the case of NCAR where the Weather Research and Forecasting (WRF) model is a next-generation mesocale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. WRF is suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometers. This is also the case in several universities in the USA and, in Europe, the Laboratoire de Météorologie Dynamique previously mentioned, but also the Max Planck Institute for Meteorology in Hamburg, Germany, and several others. Most impressive in that respect is the European Center for Medium Range Weather Forecasts (ECMWF).

Located in Reading near London, the ECMWF is a scientific center that works together with, and assists, Eumetsat and all interested weather offices or institutions in providing better service and forecasts with increasing accuracy (see Box 11.2). More than 5 million data points are processed daily in the Center and this number is continuously increasing. The exemplary association of the international scientific community, of space agencies, and of the various meteorology offices, has indeed led to the most impressive progress witnessed in weather forecasting. But this is not all! Figure 11.2 illustrates the promise of

Box 11.1 The World Meteorological Organization

Located in Geneva, the WMO is an intergovernmental organization with a membership of 188 Member States and Territories. It was established in 1950, and is the specialized agency of the United Nations for meteorology, weather and climate, operational hydrology and related geophysical sciences.

Since its establishment, WMO has played a unique role at the service of the whole planet. Under WMO leadership and within the framework of WMO programs, National Meteorological and Hydrological Services have contributed substantially to the protection of life and property against natural disasters. WMO has a unique role within the UN system, as it facilitates the free and unrestricted exchange of data and information, products and services in real or near-real time on matters relating to safety, food security, water resources and transport, economic well-being and the protection of the environment.

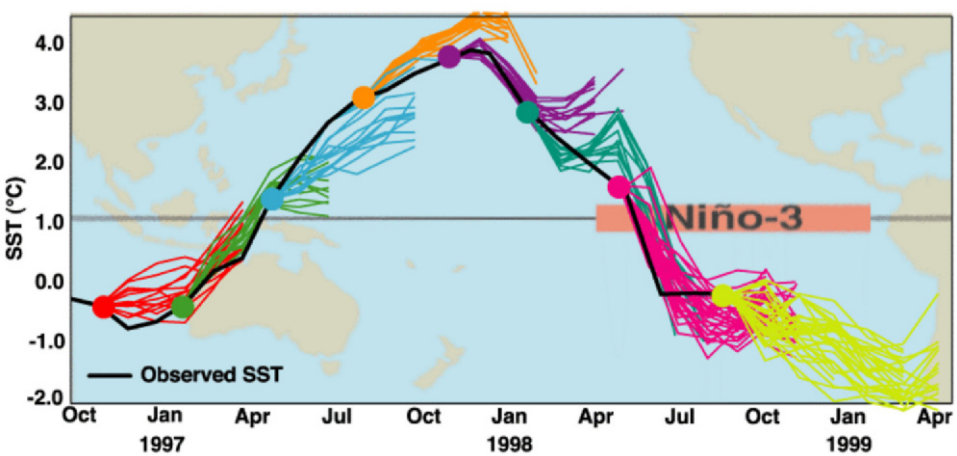


Figure 11.2 Three-monthly interval predictions and observed variations in the anomalies of sea surface temperature (SST) for the El Niño-3 area for the period 1997–1998. The first point on each of the forecast trajectories is the observed value for the month in which it started. (*Credit*: ECMWF.)

medium-term predictions over several months. The value of sea surface temperature anomalies with respect to an established model, for the range of latitudes and longitudes 5°N–5°S, 90–150°W, called Niño-3, is often used as an indicator of El Niño activity. The plots established at the ECMWF superimposed on a map of the area, show 6-month forecasts of monthly mean anomalies made every 3 months, whose real values are shown on the dark line. These plots are available on the ECMWF website [3]. The agreement is quite spectacular,

Box 11.2 The ECMWF

The ECMWF is an independent international organization established in 1975. It is supported by Belgium, Denmark, Germany, Spain, France, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Austria, Portugal, Switzerland, Finland, Sweden, Turkey, and the United Kingdom. In addition, the Czech Republic, Estonia, Iceland, Croatia, Lithuania, Hungary, Morocco, Romania, Serbia and Slovenia are participating. It has also cooperation agreements with the WMO, EUMETSAT, ESA, the African Center of Meteorological Applications for Development, the Joint Research Center of the European Union, the Preparatory Commission for the Comprehensive Nuclear Test-Ban Treaty Organization, and the Executive Body of the Convention on Long-Range Transboundary Air Pollution.

The principal objectives of the Center are:

- the development of numerical methods for medium-range weather forecasting;
- the preparation, on a regular basis, of medium-range weather forecasts for distribution to the meteorological services of the Member States;
- scientific and technical research directed at the improvement of these forecasts;
- the monitoring and storing of appropriate meteorological data.

Since June 1979, the Center has been producing operational medium-range weather forecasts. In addition: the ECMWF makes available a proportion of its computing facilities to its Member States for their research; it assists in implementing the programs of the WMO; it provides advanced training to the scientific staff of the Member States in the field of numerical weather and climate prediction; and makes the data in its extensive archives available to outside bodies.

evidencing that medium-term forecasting is offering some very promising prospects.

Forecasting the weather is certainly 'easier' than forecasting the climate or the climate-related hazards: cyclones, floods and droughts, not mentioning natural disasters such as earthquakes and volcanic eruptions, although some progress is in view here as discussed in Chapter 10. It is certainly encouraging, because of its very successful role in weather–related matters, that the WMO is now considering climate forecasting as one of its future tasks. It is also encouraging to watch the development of a broader international fleet of coordinated space missions that will help to narrow the divergent or different interpretations.

11.2.2 The scientific alert phase: the example of the IPCC

The need of a science expertise is hardly refuted by the politicians, the decision

makers and all those who are or will be key players in the future management of the planet. This is undoubtedly the case for the understanding and forecasting of global warming and climate change. Other examples will certainly gradually appear in the case of other climate-related or anthropologically relevant phenomena in the forthcoming years. Nevertheless, climate change is one of the most complex issues. Policy makers clearly need an objective source of information about the causes of the phenomenon, its potential environmental and socioeconomic consequences and the adaptation and mitigation options to respond to it.

The political perception that decisions to protect the planet from anthropogenic large-scale deterioration should be based on a sound scientific judgment probably goes back to the early 1970s. That was the time when the USA expressed strong concern that the stratosphere and the ozone layer might be damaged by the vapor trails generated by Concorde, the French–British supersonic aircraft. A substantial number of scientific studies were engaged to assess that concern. This coordinated effort, led by the United States on an international basis, had to be seen in the context of the growing competition that was developing between the United States and Europe on the development of supersonic air transportation, with the United States not easily accepting the fact that they would not have complete leadership. The studies on the damage to the atmosphere were inconclusive, and Concorde was allowed to fly between Europe and New York for more than 20 years.

Climate forecasting involves a large amount of scientific work. It is not just the prolongation of weather forecasting that deals with daily averages of data. In the case of climate predictions only the trends are meaningful. They should reproduce seasonal variations, interannual variability such as El Niño and paleoclimates. Climate depends on a variety of phenomena and parameters that evolve, suddenly becoming more important than they were previously, and these will take a long time to properly identify and forecast individually. It rests on both a large set of observations, measurements and data from the ground and from space, and on theories of physics, chemistry, mechanics and hydrodynamics. But above all, climate prediction is based on computer models (see Box 6.1 'Climate models' on page 191). Several generations of models have been developed and improved over the years, aiming at a better and more complete representation of observations and measurements: for its 2007 predictions, the IPCC used not less than 19 models! Models are based on basic physical equations, including sources and sinks, conservation of energy, momentum and water. They also include empirical parameters derived from the present climate, and are tested continuously against reliable data and from paleoclimate data. Present models have a coarse resolution of 100–200 km and are very demanding in computer power, which is probably where the bottleneck lies, as forecasts also rest on different sets of scenarios, initial conditions and assumptions such as population growth, energy consumption, the rates of emission and the concentrations of greenhouse gases (GHG). Such efforts – which involve so many variables as the Earth is treated more and more as a system and no longer

as a set of individual more or less independent layers – represent a formidable scientific undertaking that naturally necessitates the use of simplifications and lead to the establishment of margins of uncertainty. Nevertheless, systematic intercomparisons indicate an ongoing improvement. By the end of the second decade of the century, it is foreseen that century-long simulations will be achievable with a resolution of 1 km.

Of course, intrinsic uncertainties affect the predictions. These are, for example, the effect of clouds, which are crucial for the heat balance, of aerosols and of the degree of heat absorption in the oceans. Indeed, these are unavoidable and the goal of such a complex scientific effort is to reduce them as much as possible. Despite it resting on a most rigorous scientific approach, this difficult exercise leads only to probabilities and not to a mathematical certainty, and this is the source of heated discussions and even disputes. For example, when the weather forecasts indicate a 90% probability that it will rain today, you wisely dress appropriately and take an umbrella. You travel by plane because the probability of a crash is extremely small even though it is not strictly equal to zero. But you would indeed hesitate if you were invited to board a plane whose probability of crashing was 1 in 3!

Remedying, or at least slowing down, the main causes of climate changes such as the emissions of GHG, will certainly imply difficult policy decisions, a problem that we address in the following section. Uncertainties will probably open the way to a 'wait-and-see' approach, which might unnecessarily be amplified by disagreements among the scientists. Disagreements among scientists who want to promote themselves forward, do not lead to any convincing argument but rather facilitate the no-decision approach. The public releases of IPCC reports reveal these delicate issues in sharp contrast. The paradigm of what we call the alert approach is indeed the route followed by the IPCC.

The IPCC was established in 1988 to provide the decision makers and others with an objective source of information about climate change. The Panel was tasked with the preparation, based on available scientific information, of reports on all aspects relevant to climate change and its impacts on society, and to formulate realistic response strategies. The IPCC is a scientific intergovernmental body set up by the United Nations Environment Programme (UNEP) [4]. It is open to all member countries of WMO and UNEP, and its reports are based on scientific evidence that reflects existing viewpoints within the scientific community. The comprehensiveness of the scientific content is achieved through contributions from experts in all regions of the world and of all relevant disciplines, including, where appropriately documented, industry, literature, traditional practices, and a two-stage review process by experts and governments. Thousands of scientists all over the world contribute to the work of the IPCC as authors, contributors and reviewers. The IPCC does not conduct any research, nor does it monitor climate-related data or parameters. Its role is to assess on a comprehensive, objective, open and transparent basis, the latest scientific, technical and socioeconomic literature produced world wide, relevant

Box 11.3 The United Nations Framework Convention on Climate Change

The UNFCCC is establishing the framework needed at the world level to formulate the rules and define the targets aimed at the control of the environment and the climate. One of its ultimate objectives is the stabilization of greenhouse-gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system, within a time frame that would allow ecosystems to adapt naturally, to ensure that food production is not threatened, and to enable economic development to proceed in a sustainable manner.

The Conference of the Parties to UNFCCC is the highest decision-making authority of the Convention. Relevant international organizations include the IPCC and the Global Climate Observing System [5]. But many more organizations such as UNESCO, the WMO, the Food and Agriculture Organization (FAO), and the Global Environment Facility, are also involved. Similarly, many world research programs are coordinated under United Nations leadership such as, the Convention to Combat Desertification, the Development Program, the Environment Program (UNEP), the Conference on Trade and Development, etc. A more complete list is available on the United Nations website (www.un.org). One of the main problems is the efficient coordination of all the multitude of organizations, conventions and programs and their endorsement by the governments, given the varying level of respect and faithfulness that some governments grant to the United Nations! In the end, it is the governments that vote the budget and approve the programs, and their support is necessary!

to the understanding of the risk of human-induced climate change, its observed and projected impacts and options for adaptation and mitigation.

The IPCC issues its reports at regular intervals to governments, policy makers, experts, students and the public. The findings of the first Assessment Report of 1990 played a decisive role in leading to the United Nations Framework Convention on Climate Change (see Box 11.3), which was opened for signature in the 1992 United Nations Conference on Environment and Development, also called the Rio de Janeiro Earth Summit, and entered into force in 1994. It provided the overall policy framework for addressing the climate change issue. The Second Assessment Report of 1995 provided key input for the negotiations of the Kyoto Protocol in 1997. Its last and fourth report was published in January 2007 involving a group of 2,500 scientific expert reviewers from 130 countries (see also Table 6.1 on page 188). IPCC reports should, in principle, be neutral with respect to policy, although they need to deal objectively with policy-relevant scientific, technical and socioeconomic factors. They should be of a high scientific and technical standard, and aim to reflect a range of views,

expertise and a wide geographical coverage. When governments accept the IPCC reports and approve their 'Summary for Policymakers', they acknowledge the legitimacy of their scientific content.

The relationship between the UNFCCC and the IPCC has set an example and a model for interaction between science and decision makers. Because of its engagement and seriousness, it was awarded the 2007 Peace Nobel Prize shared with former Vice President Al Gore '*for their efforts to build up and disseminate greater knowledge about man-made climate change and to lay the foundations for the measures that are needed to counteract such change*'. Similar mechanisms are now considered for other environmental issues. This is the case for water resources, sustainable development and the biophysical aspects of climate change and its impact on biodiversity, for which politicians and the various economic actors also need a sound scientific expertise to assess, on a rational basis, the dangers that are faced. According to a United Nations report issued in 2000 and prepared by some 1,360 experts, about 60% of the services provided by ecosystems, and which allow humanity to survive on Earth, are already degraded or over-exploited. Consequently, an Intergovernmental Mechanism of Scientific Expertise on Biodiversity (IMoSEB) is being set up in collaboration also with the UNEP to prepare for the establishment of what could become an Intergovernmental Panel on Biodiversity Change.

11.2.3 Organizing the space tools

In all the previous chapters we have shown the significance of space data for properly understanding the phenomena potentially affecting the Earth, as well as their causes. They must, however, cope with two important requirements: first, all the needs demanded by the complex Earth system must be embraced; second, continuity must be ensured. As an illustration we use the evolution of the ozone hole above Antarctica.

We do not know precisely when the ozone hole first appeared in the history of Earth: is it a recent phenomenon, as is usually assumed, or is it a cyclic or a permanent problem? We cannot say for sure because reliable data were not available prior to 1957 when the first instruments were deployed world wide at the occasion of the International Geophysical Year, and the first global ozone measurements were made. Since then, owing in particular to the contribution of the Antarctic Survey Stations, the total ozone content has been measured. This led to the first perception that the ozone loss was due to the growth of inorganic chlorine and to how its growth might affect the ratio of NO_2 over NO in the stratosphere above the South Pole. Prior to that time, it is not possible to say for certain that the phenomenon existed. This example illustrates the necessity not only to develop the right set of instruments but also to properly calibrate and cross-compare their data and archive them for a very long time into the future. These data are a patrimony for scientists to assess the evolution of parameters that determine our physical living conditions, and must be properly preserved. Some of these data may have a commercial or strategical interest and it may not be wise to make them public. In the future, however, this restriction should be

abandoned, at least after a predetermined time, and the data be passed to the public domain as soon as possible.

Another requirement is to ensure the continuity of the data collection and of the capacity to operate coherently all the required components of the Earth observation system. This issue is critical! In several cases, ground-based components of systems are supported by research programs that do not have guarantees for long-term funding. In some cases, such as *in-situ* atmospheric networks and hydrological networks, the number of observations is decreasing because of lack of personnel to operate or maintain the stations. In other cases, the reorganization of priorities may lead to lowering the support for Earth science space missions.

That was the case at NASA when priorities were given to manned exploration of the Moon and of Mars. As a result, aging satellites, several of which had already gone beyond their planned lifetimes, were not systematically replaced in spite of their essential role [6]. The workhorses of operational Earth observation satellites in the United States, the Landsat series, will face a crisis and a gap in data collection if the Landsat Data Continuity Mission (LDCM) is not successfully launched in 2011. Equally worrisome is the situation of weather monitoring in the United States: as the costs of the National Polar-Orbiting Operational Environmental Satellite System (NPOESS) keep growing, the funding for the instruments to fly on these satellites has been cut [7]. This illustrates the need to go a step further than the scientific research level and to bring to the front the strategic character of the problem.

In the early days of the space era, a substantial number of Earth observation missions were launched, but the need to coordinate them was not felt at that time to be critical. Also, the issue of combining their data for conducting a systematic study had not at that time been perceived as critical: at the onset of the space venture, everything was important and it did not matter very much whether Earth missions were part of a coherent program or not. The decisions to develop them were reflected in the scientific and the political interests of the institutes, agencies or ministries involved. The establishment of some kind of a road map addressing the most pressing needs was not considered to be urgent and it was *de facto* relegated to some undefined future. Little by little, however, the situation started to change as science knowledge accumulated, giving sharper evidence of the need for a system approach.

The space-faring nations started to realize the challenge with which they were confronted if they wished to make intelligent usage of the satellites they were developing in such impressive numbers. In 1984 they created the Committee on Earth Observation Satellites (CEOS) that attempted to coordinate all civil space-borne missions devoted to the Earth [2, 8]. CEOS included some 26 members, most of them being space agencies, along with another 20 associated national or international organizations, totaling about 170 satellites, carrying over 340 different instruments, to be operated over 15 years. CEOS is now recognized as the major international forum for the coordination of Earth observation satellite programs and for interaction of these programs with users of satellite data world wide.

CEOS followed and paralleled a similar group of agencies established in 1981 among a more reduced set of four: i.e. NASA, ESA, ISAS (the Japanese Institute for Space Astronomy and Science) and IKI (the Institute for Cosmic Studies of the Russian Academy of Sciences). This group, named the Inter Agency Consultative Group (IACG) was set up by the four space organizations to coordinate their six respective missions aiming at the study of Halley's Comet at the occasion of its passage at perihelion in 1986 [9]. The IACG proved to be very useful for improving the science return of each respective mission. Its mandate was prolonged after 1986 to coordinate missions of the same agencies in the field of Solar and Terrestrial Physics. It was then left dormant and abandoned mostly because of a lack of sharp focus just after celebrating its 20th anniversary in 2002. The IACG can be considered nevertheless as an excellent example of the willingness of different agencies to work together for the benefit of science, taking advantage of a broad international cooperation framework increasing the scientific return through the acquisition of a maximum number of measurements and observations.

The success of CEOS was more difficult to achieve. It was not so obvious at the start that it might become useful and not just remain a kind of bureaucratic forum. These initial difficulties can be explained by the relatively broad spectrum of Earth observation missions as compared to only six in the case of the IACG, and by the unavoidable interference of substantial political interests from the participating agencies overriding the pure scientific matters. Nevertheless, after some 20 years of existence, CEOS can reasonably be considered to bear some fruits, owing to a better and more efficient organization among all the participants. The sustained understanding by the space agencies – that it is useful to have such a tool – will from now on hopefully ensure the provision of the data necessary for a better management of the Earth system through space missions. CEOS can be considered as a first step towards the development of a more complete space system and the lessons learnt from its early stages can help to develop the next and more efficient steps that will be required in the future.

The need to coordinate different satellites into a system is now extending to other and somewhat more focused grouping of agencies, such as the Global Observing System (GOS) of WMO. GOS (Figure 11.3) provides ground-based and space observations of the state of the atmosphere and of the ocean surface for the preparation of weather analyses, forecasts, advisories and warnings, for climate monitoring and environmental activities carried out under WMO and other relevant organizations. It is operated by National Meteorological Services, space agencies, and involves several consortia dealing with specific observing systems and five geographic regions covering the east and west Atlantic, the east and west Pacific and the Indian Ocean. About 30 satellites from the United States, Russia, Europe, Japan, India, China and South Korea are involved in the program which was established in 1990. The GOS is the most important program of WMO for observing, recording and reporting on the weather, climate, and the related natural environment, and for the preparation of operational forecasts, warning services, and related information.

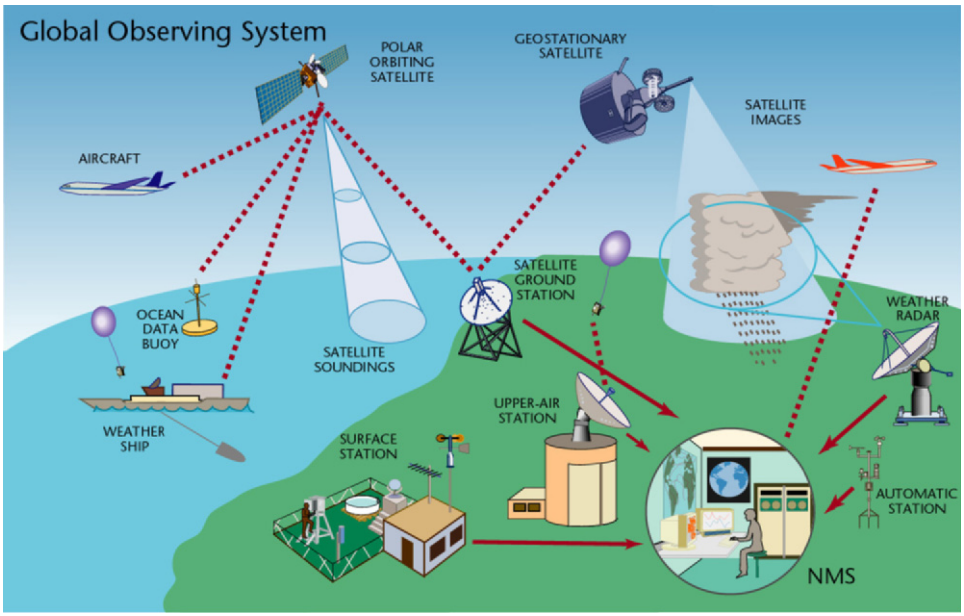


Figure 11.3 The concept of the Global Observing System of the WMO, involving space- and ground-based tools. (*Credit*: WMO.)

WMO is also addressing the important issue of intercalibrating satellite measurements through a Global Space-based Inter-Calibration System asking the responsible space agencies to take steps to ensure better comparability of different measurements and tying those to absolute references. This is an essential requirement to properly track climate change as, in the past, flawed data have caused disagreements between the scientists and seriously disrupted the study of climate trends [10, 11].

In Europe, in view of the apparently increasing amplitude of natural catastrophes, ESA and CNES, the French space agency, decided to join forces and make the access to space technologies easier for potentially vulnerable countries through the establishment of the International Charter on Space and Major Disasters. This initiative aims at reinforcing international cooperation in humanitarian help by improving the efficiency of emergency services and usage of space observation, telecommunication and meteorology satellites to assist the victims of natural disasters.

Other examples of successful coordination include the Disaster Monitoring Constellation (DMC) coordinating micro-satellites from Algeria, Turkey, Nigeria, China, Vietnam, Kenya and the UK! These imaging satellites have a relatively modest resolution of only about 30 meters, which is sufficient, however, to determine the extent of areas affected by natural disasters or for the monitoring of land and cultures [12]. It is foreseen that soon they might reach 2.5 to 5 meters. They operate as a constellation, which allows them to revisit the same

area every day, avoiding cloud coverage and allowing rapid data acquisition. Even though this consortium has a commercial vocation, they offer a percentage of their data for free to humanitarian organizations.

Another constellation involving the United States, Japan and Europe, is centered on NASA's Global Precipitation Measurement (GPM) mission. Linked by Internet, the elements of this constellation allow a better understanding of global weather patterns and improved local forecasting for storm-watching, for example, providing early warning of the intensity and the paths of hurricanes, of impending floods and landslides. The power of such a system was demonstrated at the occasion of Katrina and Rita hurricanes in 2005 using radar and passive sensor measurements data from the joint US–Japanese Tropical Rainfall Measurement Mission (TRMM), and could analyze the structure and intensity of both hurricanes.

A new constellation dubbed COSMIC (Constellation Observing System for Meteorology, Ionosphere and Climate), under the responsibility of Taiwan, and the University Corporation for Atmospheric Research in the United States, has six micro-satellites tracking tiny changes in the speed of GPS radio signals and can provide vertical profiles of temperature and water vapor at more than 1,000 points every day. With COSMIC data, stratospheric temperature forecasts over the northern hemisphere have significantly improved, and it seems possible also to predict the occurrence of cyclones and to achieve long-range weather forecasts. COSMIC also measures electron density in the ionosphere, an important observation for forecasting Sun–Earth connected phenomena (see Chapter 10).

In the future, several hundreds of satellites will be necessary to properly address the different aspects of climate and environmental change. Others will be employed to assist decision making in the strategic planning and management of industrial, economic, and natural resources, as well as in the provision of information required for satisfying a sustainable development. However, satellites alone cannot solve all Earth's problems and they must be complemented by measurements made from the ground. Ground-based stations must therefore be integrated into the whole system of measurement, of warning and protection. As new operational services are being established in different domains, the need for creating an operational system, including all key elements and providing the indispensable data to the competent scientific and operational services, becomes more necessary. How can we imagine ultimately coordinating the whole set of nations that share the planet and its resources if we cannot coordinate technical systems and scientific research which, in principle, is a much simpler task? This century, if not the next decades, should prove that this is achievable. The setting-up of the Group on Earth Observation (GEO) and the definition of the Global Earth Observation System of Systems (GEOSS) described in Box 1.4 on page 391 are certainly encouraging signals that things are moving in the right direction.

11.3 The indispensable political involvement

Managing the Earth's future is clearly a global issue. It concerns first the scientific world where international cooperation is the rule and is therefore relatively easy to involve. It concerns also the nations and their societies which, of course, are not global and, as seen daily in the present times, are much more difficult to coordinate because of their historical differences, their different economic interests or their different approaches to political systems, not mentioning religion. Managing the control of climate change and of the global environment requires, first, a political advanced perception, following the scientific evidence that the issues have reached a critical degree and, second, decisions that must be taken at the world level in anticipation of their possibly catastrophic consequences. Nations, however, have a long life time and are often reluctant to change, are seemingly incapable of thinking about the future, and prefer not to organize responsibly and implement the necessary re-orientations of their policies or behaviors. In that respect, three countries will, by their attitude, have a decisive role on the future of the planet.

11.3.1 The crucial role of the United States, China and India

In the forthcoming years, the United States and China will bear an enormous responsibility in the preservation of our future living conditions. With 24% of the world's total, growing at the rate of 1.5% per year, the USA is the main contributor to carbon emission, far above any other nation with some 50% of its electricity being produced from carbon burning, making the USA presently the biggest polluter in the world, as shown in Table 11.1. With a share of nearly 40%, the USA is the largest contributor in the world of GHG emissions. Such a prominent position largely explains the US Administration's reaction to the Kyoto Protocol, an issue that we develop in the next section. It is clear that any global policy must take the US position into serious consideration and, reciprocally, it is useless to define a policy that the USA would not adhere to.

Table 11.1 CO_2 emissions in 2005 and projected to 2030 for the world (population of 8.2 billion assumed according to the medium variant of the UN predictions) and six main industrialized countries, including the 27 member states of the European Union (adapted from the International Energy Agency and Le Monde, 8 November 2007)

2005	Gt	Tons per capita	2030	Gt	Tons per capita
1. Unites States	5.8	19.3	1. China	11.4	7.8
2. China	5.1	3.9	2. United States	6.9	18.8
3. Europe	4.4	8.5	3. Europe	4.0	8.1
4. Russia	1.5	10.7	4. India	3.3	2.2
5. Japan	1.2	9.4	5. Russia	2.0	16.1
6. India	1.1	1.0	6. Japan	1.2	10.2
7. World	27	4.2	7. World	43	5.2

China has already overtaken the United States to become the biggest emitter. It is also the largest contributor of chlorofluorocarbons and sulfuric oxides, producing twice as much SO_2 as the USA. The size of its population and the rate of growth of its economy – the fastest of any major nation – give China a special place in the present context and even more so in the future. Its demands for oil will more than quadruple as the number of its vehicles will reach more than 270 million by 2030. Coal will be the main source of energy, to a ratio of 75%, for feeding its power plants. The huge needs for energy of China and India represent one of the main challenges for the planet in the decades to come. However, Table 11.1 shows that, per capita, China ranks far behind the USA – and below the world average on what concerns resource consumption and waste output – because of the growth of its population, which has more than doubled over the past half century, representing now more than 1.3 billion people or 20% of the world's total, and is projected to be about 1.5 billion by 2030 according to the medium variant estimates of the United Nations.

One environmental effect of the booming Chinese activity is the emission of nitrogen dioxide, NO_2 in the atmosphere (see Figure 10.27). While NO_2 concentrations tend to diminish over western and eastern Europe as well as over the United States owing to the use of more efficient technologies, they are increasing significantly in China (Figure 11.4) [13]. Direct satellite measurements show that the concentration of NO_2 in the atmosphere over China has risen by 50% in 10 years since 1995 and this is just the beginning of an accelerating process. Naturally, China therefore faces large environmental problems. Its air is polluted and it suffers soil erosion on 19% of its land area in addition to frequent floods and extensive natural disasters. Water quality is poor and declining; about 75% of its lakes are polluted as well as almost all coastal seas. Hopefully, the Chinese public's awareness on environmental issues is rising. Reforestation has started and better education is already showing its effect. The country is improving its energy efficiency at a fast pace, resulting in minimal increases in greenhouse gas emissions. The problem for the government is to be able to win the race between accelerating environmental damage and accelerating environmental protection. After the economical miracle, there is a need for an environmental miracle, allowing the Chinese people to reach socioeconomic and environmental sustainability.

For similar reasons India is the next nation to play a key role in the trend of climate change, as its energy needs will more than double between 2005 and 2030. India, which in 2005 was the last in the list of Table 11.1, will be number 4 in 2030 just behind Europe, having tripled its emissions. Its rapidly growing population, along with a move towards urbanization and industrialization, has placed significant pressure on its infrastructure and its natural resources. India's booming cities are causing serious air pollution problems. However, the Indian government is taking the environment question very seriously. The country has a separate government ministry exclusively for non-conventional energy sources, and it has one of the largest national programs to promote the use of solar as well as wind-generated energy. As urbanization picks up pace and vehicle

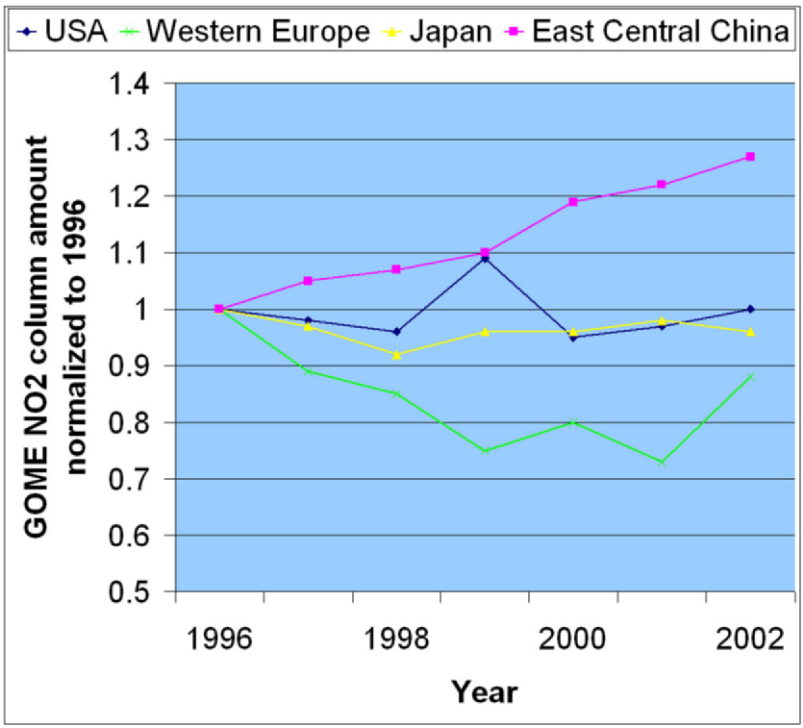


Figure 11.4 Time evolution of tropospheric NO_2 columns normalized to 1996 for selected areas as measured by the GOME instrument on board ESA ERS 2 satellite, (Adapted from reference [13].)

ownership increases, the Indian government's ability to safeguard the country's environment will depend on its success in promoting policies that keep the economy growing while providing adequate energy to satisfy its people's consumption requirements in a sustainable manner.

Certainly, both China and India have all the rights to justify their enormous energy demands and power consumption as this '*will contribute to a real improvement in the quality of life of nearly one third of the world population. This is a legitimate aspiration that needs to be accommodated and supported by the rest of the world*' says Nobuo Tanaka, Executive Director of the International Energy Agency [14]. Nevertheless, if governments around the world do not change their present policies, the world energy demands will be more than 50% higher in 2030 than in 2007, with China and India together accounting for 45% of the increase. These three examples illustrate better than any other the absolute necessity of a political engagement of all governments around the world in addressing the stabilization and the future management of the planet as their economical status and growth will determine the balance of all resources on Earth, of the environment and the well-being of all.

11.3.2 A perspective view on the political perception

The concept that the future of the planet must be secured through proper political industrial and ecological management and that its sustainable development must be ensured, appeared for the first time at a Conference held in Stockholm in 1972 but it was not until 1987 that a definition of that term was found, in the Brundtland Report (see Chapter 1): '*a development that responds to the present needs without compromising the capacity for the future generations to respond to theirs.*' In June 1992, the Earth Summit of Rio marked a first step towards the establishment of worldwide governance with the signature of a series of international conventions on climate, biodiversity and desertification. The Rio summit generated some optimism. In the same year, the UNFCCC (see Box 11.3 on page 375) committed most industrialized countries to return their emissions of GHG to the 1990 levels by 2000. That also marked a genuine progress. Unfortunately, the slowness of the subsequent negotiations failed to lead to any concrete substance. Important steps, however, were made with the establishment of the Montreal and Kyoto Protocols.

The Montreal and Kyoto Protocols

The 1985 Vienna Convention for the Protection of the Ozone Layer which outlines states' responsibilities for protecting human health and the environment against the adverse effects of ozone depletion, established the framework under which the Montreal Protocol was negotiated. The Montreal Protocol on substances that deplete the ozone layer is a landmark international agreement designed to protect the stratospheric ozone layer. It stipulates that 'The production and consumption of compounds that deplete ozone in the stratosphere – chlorofluorocarbons (CFCs), halons, carbon tetrachloride, and methyl chloroform – are to be phased out by 2000 (2005 for methyl chloroform)'. The treaty was originally signed in 1987 and came into force in 1989. It has been ratified by more than 180 countries while being substantially amended four times, in particular in 1990 in London and 1992 in Copenhagen, as the scientific basis of ozone depletion became available for the principal halogen source gases. Further amendments were agreed in Vienna (1995), Montreal (1997) and Beijing (1999). The predicted variations of abundances of effective stratospheric chlorine, which also includes bromine, are shown on Figure 11.5 [15] for different scenarios: without the Protocol and consequent to its various amendments. The effect of the Protocol is impressive, since without it, the stratospheric halogen gases were projected to have increased significantly in the 21st century.

The next question is, of course, the effect on the ozone layer. Models, which are used to assess past changes and predict future changes, are built on different assumptions concerning atmospheric chemical composition and rate of climate change. The present situation and its projected evolution are shown on Figure 11.6 [16]. The ozone layer is expected to recover by the middle of the 21st century, assuming global compliance with the Montreal Protocol. The range of model predictions indicated by the blue shaded zone shows that the lowest

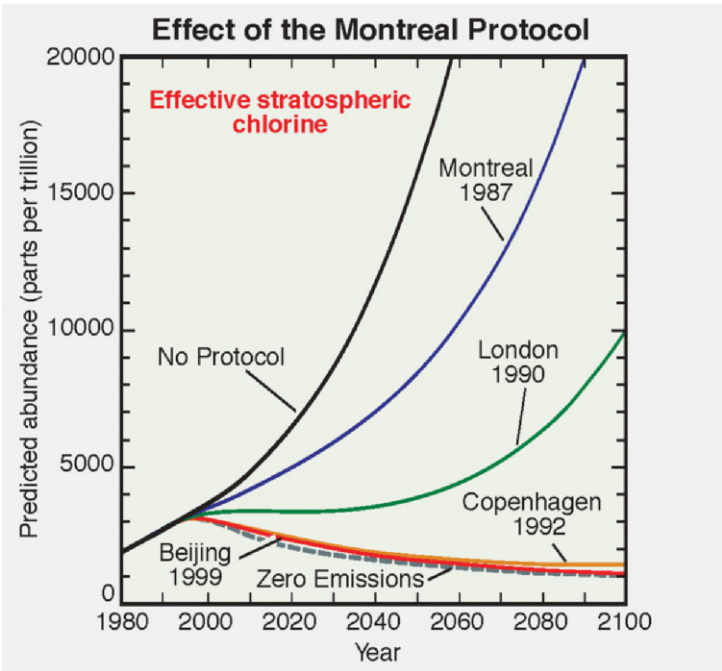


Figure 11.5 The projected effects on effective stratospheric chlorine abundances are shown in the various cases defined by the Montreal Protocol of 1987 and by its different amendments. The 'Zero emissions' line assumes that all emissions are reduced to zero in 2003. (Adapted from reference [15].)

ozone values are expected to occur about 2008/2009. However, even though the Protocol seems to have had a positive effect on the reduction of CFCs, the future outlook for ozone recovery is uncertain. Volcanic eruptions could delay the recovery process, as shown by the example of the Pinatubo (Figure 10.5), and global warming may accelerate or delay ozone recovery depending upon actions taken on GHGs. What these models put clearly in evidence is the relatively long time it takes the ozone layer to return to its pre-1980 situation after the Montreal measures have been implemented. This illustrates the need to act as soon as possible and stop the anthropogenic alterations of the atmosphere. This is particularly urgent in the case of the emissions of GHGs, as stated by the Kyoto Protocol that we now address.

The Kyoto Protocol is an amendment to the UNFCCC. Countries which ratify this Protocol commit to reduce their emissions of carbon dioxide and five other GHGs by a collective average of 5% below their 1990 levels by 2008–2012. While the targets for reduction is on average 5%, national limitations range from an average of 8% across the European Union, to 7% for the USA, 6% for Japan, 0% for Russia, with permitted increases of 8% for Australia and 10% for Iceland. Countries do not need to sign the Protocol in order to ratify it; signing is a

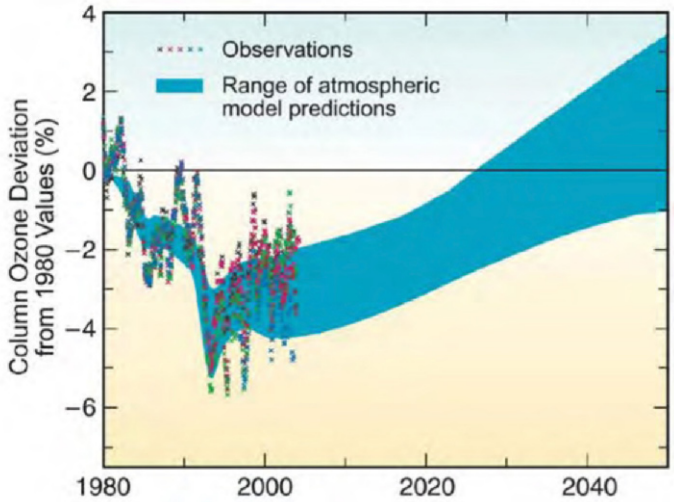


Figure 11.6 Observed and modeled column ozone amounts (60°S–60°N) as percentages of deviations from the 1980 values. *(Source*: IPCC [16]).

symbolic act only. Conversely, the Protocol is non-binding unless ratified. A distinction exists between developed nations and developing nations such as China, India and Brazil. The latter were exempted of targets, because they were not the main contributors to climate change at that time. This was one of the reasons behind the United States opposition to the Protocol. China has since then ratified the Protocol and is expected to make herself no longer exempt.

The Protocol was negotiated in Kyoto in December 1997, hence its name, and came into force on 16 Feburary, 2005 following ratification by Russia on 18 November, 2004. A total of 175 countries have ratified the agreement, representing 61% of emissions from developed countries. One notable exception is the United States (and Kazakhstan!), which saw Kyoto as a scheme to either retard the growth of their industry or to transfer wealth to the third world in what they claimed to be a 'global socialism initiative'! We shall dwell on the US resistance in the following section.

The capping imposed on the ratifying countries is associated with possible emission trading between them through various shared 'clean energy' programs and 'carbon dioxide sinks', such as forest replanting and underground storage. For example, industrialized countries can trade emission quotas with developing countries providing they invest in these countries through projects that would contribute reductions in GHG emissions in the form of forests and other systems that remove carbon dioxide from the atmosphere. Another mechanism allows industrialized countries to collaborate, such as Germany funding a clean energy project in Russia. In that case, Germany would have been credited a reduction of its own quota.

Some estimates indicated that if successfully and completely implemented,

the Protocol should reduce the average global rise in temperature somewhere between 0.02°C and 0.28°C by the year 2050, which should be compared with an increase of 2°C to 4°C between now and 2100, as predicted by the IPCC at its fourth meeting in January 2007, which is evidence of its far too modest ambition [17]. Because of this, some environmentalists questioned the value of the Kyoto Protocol, insisting on imposing tougher commitments and new measures, should the current ones fail to produce deeper cuts in the future. In a move to avoid implementing the measures, others used the poor performance to claim that Kyoto did not offer the proper framework to achieve the goals of stopping global warming. Among them the United States have been the most resistant.

The United States resistance

Under the Clinton Administration, the President himself, concerned that the Senate may not accept the terms, never submitted the Protocol for ratification. Nor did George W. Bush submit the treaty for ratification on the basis that it would put the US economy under unacceptable strain and because of the uncertainties that were present in the whole climate change issue [18]. A study made under the Department of Energy estimated that the United States would have to reduce its annual carbon emissions by about 540 million tons between 2008 and 2012, equivalent to turning off about 90 coal power stations each year, costing the economy around $400 billion.

The prospect of the USA staying outside the Kyoto agreement influenced a number of other countries such as Australia, Japan and Canada, which saw Kyoto as a means to put them at a competitive disadvantage with the United States. While Japan ultimately decided to ratify the Protocol, Prime Minister John Howard of Australia, one of the biggest emitters of CO_2 per capita, said he would not and would look for 'fairer' alternatives that would have no mandatory controls on GHG emissions of some countries and not of others whose economies were booming. He was, of course, referring to China and India, which did not have any obligations. In November 2007 a new government was formed by the Australian Labor Party under Kevin Rudd as new Prime Minister, who fully supported the Protocol and ratified it immediately at the eve of the United Nations Conference in Bali. In January 2006, under the leadership of Stephen Harper and his Conservative minority government, Canada, which had first ratified the Protocol at the end of 2002, reverted to a more protective position aligned on that of the United States. Nevertheless, and regardless of the national anti-Kyoto position, some individual provinces, including Quebec, British Colombia and Manitoba, were pursuing Kyoto-type measures.

One of the reasons behind the United States anti-Kyoto attitude was their opposition to accept that an international consortium of nations orchestrated under the banner of the United Nations, would dictate policies that clearly undermined the USA's own leadership. The concept of leadership is deeply anchored in the way of thinking of the successive US administrations and is certainly not unpopular in the country, explaining the very protective attitude

adopted in the different international meetings where the Kyoto Protocol and its successor were addressed. Leadership can certainly explain many of the initiatives taken by the government against the Kyoto Protocol in organizing diversion approaches. At the meeting of the Association of South-East Asian Nations (ASEAN), on 28 July 2005, both the United States and Australia proposed a 'fairer alternative', the Asia Pacific Partnership on Clean Development and Climate, an agreement between six Asia-Pacific nations: Australia, the People's Republic of China, India, Japan, South Korea, and the USA. The pact allowed those countries to set individual goals for reducing GHG emissions, but with no enforcement mechanism. Supporters of the pact saw it as complementing the Kyoto Protocol while being more flexible. Critics and environmentalists saw it as a scheme to wreck the Kyoto Protocol, adding that the pact will be ineffective without any enforcement measures [19].

Similar initiatives were orchestrated in May 2007 just before the G8 summit in Heiligendamm in Germany, always trying to avoid the imposition of binding commitments. Sentences such as '*When we burn fossil fuels, we emit GHG in the atmosphere . . .*' were considered to reflect an important way forward and an opening in the position of the United States! Hiding behind the intrinsic scientific uncertainties, in spite of the continuously increasing probabilities of the reality of global warming and of its foreseen devastating consequences as described and substantiated in the 2007 IPCC reports, they could not do less than agree that human activities '*contribute in large part to increases in GHG'*. But no measure was initiated to seriously curb down the temperature rise of the planet. At the UNFCCC meeting in Montreal in December 2006, when the extension of Kyoto beyond 2012 was approved by the 189 participating nations – who also agreed to take parts in negotiations aiming at deeper cuts after 2012 – the United States again opposed such commitments. They proposed instead to engage in technological innovations and in partnerships with smaller countries. They also accepted to remain members of the Convention which, contrary to Kyoto, is not binding, allowing some kind of dialogue to continue with other nations. They invited developing countries to be part of the future discussions on medium and long-term target reductions – a move considered essential in their view, given the booming development of these countries. In his State of the Union Address in January 2007, the President announced a plan to develop new technologies '*that will help us to confront the serious challenge of climate change'*. Indeed, such 'official' statements were never expressed so clearly before and in evaluating the accomplished progress, it could be interpreted as an irreversible step and another positive sign that the President of the United States was slowly moving in a more positive direction.

Europe's ambitions and contradictions

Europe has visibly adopted a leading position of principle as to what concerns environmental and civil security issues with the adoption of the GMES program (see following section), not only at the level of the European Commission, but also of individual countries and other European entities such as ESA. For

example, European Union environment ministers have proposed that in compliance with the IPCC recommendations preventing the Earth's temperature from rising above 2°C before 2100 – the commonly quoted threshold for dangerous effects of climate change – the developed nations should reduce their emissions by 30% in 2020 relative to 1990, a reduction up to six times as severe as Kyoto, through in particular a combination of regulatory and technological measures. They asserted that recommendation, however, with the condition that all industrial countries adopt similar targets and that the most advanced developing nations also contribute to that goal in accordance with their responsibilities (in global warming) and their respective possibilities. In the absence of such an agreement, the European Union would nevertheless adopt a target of 20%. In Bali, in December 2007, they went even further by proposing a target of 60 to 80% for 2050.

At the level of individual nations, the UK was also taking advanced positions, announcing that it would slash its GHG emissions to 60% of its 1990 levels by 2050, far beyond the Kyoto targets, a program associated with cuts of a much longer term than any of the major polluting countries had adopted so far. Critics pointed out that if the plan was ambitious in the long term, it was not tough enough in the short term. In October 2006, the UK government published the report prepared by N. Stern, Doctor of Economy at Oxford and a former chief economist at the World Bank, which had a considerable impact in Europe, if not in the United States. The report concluded that tackling Climate Change would cost 20 times less than doing nothing, in sharp contradiction with the oil and coal industries and with the United States President claiming that the Kyoto targets would cost US$ 400 billion until 2012 [20].

In parallel, France announced a targeted reduction of a factor of 4 by 2050, without however indicating how it would practically manage to reach this target. Immediately after he was elected, French President Nicolas Sarkozy organized in October 2007 a '*Grenelle de l'Environnement*', a term that describes a forum involving government officials, experts in environmental issues and ecologists. With about 80% of its energy provided from nuclear plants, France is one of the best pupils of all nations in Europe as far as CO_2 emissions are concerned. Unsurprisingly, the ecologists at the 'Grenelle' meeting expressed their opposition to the recourse to that form of energy. The meeting was inconclusive on the matter and somewhat incoherent. In the middle of the Bali Conference, Germany, which claimed to be the champion of the ecological revolution, fixed a very ambitious 36% reduction target for 2020, resting on the use of renewable energies and on economies! In reality, Germany had a much less advanced position when it came to the point of forcing its automobile industry to develop new models emitting less CO_2, or re-enforcing the speed limits on its freeways, favoring new coal power plants and totally eliminating nuclear energy by 2021!

The European Union on its side was not providing evidence that it might meet its ambitious reduction goals. It was facing difficulties in sticking to the 1997 8% reduction targets by 2012, as it had not even been able to reach more than 2% in 2005! As in other parts of the world, environmentalists and chiefs of industries

were engaged in ferocious discussions. The former adopting a most stringent position on the limitations, and the latter trying to limit the effects of the targets on their competitiveness *vis-à-vis* other nations much less worried about the impact of their activities on the environment. Several European countries, such as Italy, Spain, Austria and Finland, were also having trouble to meet their Kyoto targets.

Europe's advanced position: GMES

In parallel, however, an impressive process reflecting a genuine European political intention was initiated: the Global Monitoring Environment and Security (GMES) program [21]. The desire to possess an autonomous and independent capacity of observation and surveillance was very strongly behind the initiation of GMES for Europe to monitor its environment and re-enforce its role on the international scene. GMES is the response to the need by Europe for geospatial information services. It provides autonomous and independent access to environment and security information for policy makers. GMES was initiated by the European Union in 1988 and was confirmed as the Union's priority at the 2001 Summit in Gothenburg, where the Heads of State and Government requested that *'the Community contribute to establishing by 2008 a European capacity for Global Monitoring for Environment and Security'*, with the aim of establishing a fully operational system between 2013 and 2015.

GMES aims at preventing natural and any other kind of disasters, as well as surveying climate change, taking care of the long-term preservation of natural resources, through the coordination of the set of data already obtained – or to be obtained – from space and from ground-based instruments. The 2 to 4 billion euros GMES, together with Galileo**,** is one of the very visible space involvements of the European Commission jointly funded by ESA. One of the objectives of that program is also to assure continuity of the space segment and to identify the successors or the complements of big ESA missions, of nationally funded projects and of the weather satellites under the responsibility of Eumetsat**,** totalizing some 30 satellites. A series of dedicated smaller missions operated around the clock, called *sentinels*, would be devoted to monitoring water resources, ocean and coastal zones surveillance, soil management, etc., using radar and optical as well as spectroscopy techniques (see Chapter 10). The idea is that by the mid-2020s, Europe should possess a system in all areas of environment monitoring equivalent to what exists in meteorology.

GMES also represents the European contribution to the international Global Earth Observation System of Systems (GEOSS: see below and Box 11.4). Interestingly, GMES is considered as a good model and a step in the right direction by American scientists who would like to see a similar approach expanded in the country and, why not, on a more international basis, paving the way to a global information and monitoring system [22]. In our view, this is one of the most immediate challenges for the nations of the world to agree on, as discussed in the rest of this chapter.

A step forward: the creation of GEOSS

What had been lacking for a long time, as discussed in Section 11.2 was a truly global organization involving all countries, the rich and the poor, that would allow the transfer and use of Earth observation data and information by all, similar to the system that already exists in the area of weather observations. In other words: a system able to instantaneously hook-up all Earth observation satellites and ground-based stations, buoys and all oceanographic and atmospheric instruments.

Box 11.4 The Global Earth Observation System of Systems

GEOSS has two main roles:

1. Make all existing systems in Earth observation more efficient through a proper coordination of all existing systems and organizations
2. Establish an exhaustive list of all the requirements necessary to create the optimum system that would ensure the future of the Earth and of its populations with the maximum degree of safety.

The second role is certainly one of the most important and fundamental, but certainly not the easiest. The ad hoc GEO 10-year Implementation Plan reference document [1] gives an exhaustive list of 10 areas where requirements and road maps have to be established, from natural and human-induced disasters, human health and well-being, energy resource management, climate variability and change, water cycle, weather information forecasting and warning, management and protection of terrestrial, coastal and marine ecosystems, support of a sustainable agriculture and biodiversity conservation, and common observations and data utilization.

GEOSS should include components consisting of both existing and future Earth observation systems, from primary observation to data and processed information production. The already existing systems would remain as such but would be supplemented by their involvement in GEOSS. Through GEOSS, they will share observations and products and make it possible to ensure that these are accessible, comparable, calibrated and responding to the users' needs. GEOSS will also attempt to identify gaps and unnecessary duplications and ensure the necessary continuity of the various components. GEOSS has a global vision and aspires to involve all countries of the world, and to cover *in-situ* as well as airborne and space-based observations.

Scientists, agencies and policy makers have been discussing for years the concept of a Global Earth Observation System of Systems**.** More concrete plans began to take shape at the Johannesburg World Summit on Sustainable Development organized in the summer of 2002 under the aegis of the United Nations, which brought together tens of thousands of participants, including

Heads of State and Government, national delegates and leaders from non-governmental organizations, businesses and other major groups. The summit was to focus the world's attention and direct action towards meeting the challenges of improving people's lives and conserving natural resources in view of the growing population, with ever-increasing demands for food, water, energy, shelter, sanitation, health services and economic security. Following this meeting and the Evian G8 summit (see below), the United States took the initiative of convening a first Earth Observation Summit in Washington in July 2003 which was attended by high-level officials from 33 countries plus the European Commission and 21 international organizations. The organization of the Washington summit was considered by some as another manifestation of defiance by the USA *vis-à-vis* the United Nations, which would normally be in charge of organizing such a meeting.

The summit adopted a Declaration aimed at developing a comprehensive, coordinated, and sustained Earth observation *system of systems*. The summit also established an ad hoc intergovernmental Group on Earth Observations (GEO), co-chaired by the European Commission, Japan, South Africa, and the United States, and tasked with the development of an initial 10-Year Implementation Plan. At the second Earth Observation Summit in Tokyo in April 2004, a Framework Document defining the scope and intent of GEOSS was adopted by 43 countries and by the European Commission, joined by 25 international organizations. The Third Earth Observation Summit meeting in Brussels on 16 February 2005 decided to transform the ad hoc GEO into the intergovernmental Group on Earth Observations presently hosted by the WMO in Geneva, charging it to take the necessary steps to implement GEOSS [1]. They '*encouraged the governments of all United Nations member states to become members of the GEO and invited the governing bodies of the United Nations specialized agencies and programs as well as all other relevant international and regional organizations to endorse the implementation of GEOSS and to encourage and assist the GEO in its work*'.

Some countries were hesitant to join in because some of their data were considered to be of a classified nature. Others felt that they could lose their independence and self-sufficiency if they shared too much of their data [23]. For example, even though India has a strong interest in the success of GEOSS, it is reluctant to share data in real time from its network of seismometers, because they are said to be vital for its national security and are held indefinitely hidden as they may pertain to nuclear testing. Commercial interests may also impede free exchange of data because some are only available for sale. Hopefully, the USA seemed very much dedicated to make GEOSS a complete success – as long as it could control the system – but could not guarantee that it could maintain its support to the initiative in the long-term as federal budgets are voted year after year and a permanently positive vote could not be guaranteed. If properly managed however, GEOSS has the potential to be one possible option for what is needed to support the alert phase and ultimately securing the future of the Earth and its increasingly vulnerable societies. It is a first and important step towards the establishment of a planetary organization, paving the way to a world scale GMES!

Trying to achieve consensus: political summits

In parallel to the United Nations initiatives, in particular the Montreal and Kyoto Protocols, the issues of global warming, of the safeguard of the planet and the need to engage in sustainable development were raised at the highest political level through several bilateral visits or multilateral meetings. Such meetings were considered essential for adopting consensus at the talks on climate under the aegis of the United Nations. Some nations also felt more comfortable being involved in smaller groups of discussion rather than under the world spotlights of the United Nations conferences. Regular occasions were offered by the G8 summits of the Heads of State of the richest countries of the world, plus Russia, meeting regularly every year in different parts of the world.

In June 2003 in Evian, France, the G8 summit affirmed the importance of Earth Observation as a priority activity and stressed the importance of GEOSS. Two years later, at the July 2005 summit at Gleneagles in Scotland, climate change was also high on the agenda. The UK Prime Minister was fully in support of the European position, and very much in favor of implementing the Kyoto Protocol. The US President made it clear that any direct reference to the Kyoto Protocol in the final resolution would force him to refuse any compromise. The outcome of the debate was unsurprisingly rather mixed. On the other hand, the G8 promised to start a dialogue with China, India, South Africa, Mexico and Brazil, in view of encouraging investment and usage of clean technologies. While acknowledging that '*climate change is a serious and long-term challenge that has the potential to affect every part of the planet*', they expressed their desire for '*practical commitments industrialized countries can meet without damaging their economies*'. In the 2007 Heiligendamm summit in Germany, energy and climate issues were again at the core of the discussions, but while Europe, with the voice of Angela Merkel, was firmly resting on the imposed targets of Kyoto and on the necessity for the richest countries to cut by a factor of 2 their emissions in 2050 in view of limiting global warming to 2°C, the USA's position was more *bottom up*, leaving the initiative of making the necessary efforts to the governments rather than respecting an international treaty. Hopes were expressed that the United Nations Conference in Bali the following December would clearly identify a robust and effective follow-up to the Kyoto Protocol! We will analyze the outcome of the Conference in the following section.

11.3.3 The emotional perception: the scene is moving

The tricks of nature and the reactions in the United States

In the meantime, the Earth was playing unusual tricks, behaving in a 'not politically correct' manner and exerting an unexpected pressure, putting in sharper focus the necessity to act. The 26 December 2004 tsunami in Asia, observed in real time on all the television screens of the world, came as a shock, as many people could be seen dying in front of TV cameras without any chance of being rescued from the violence of the waters. The countries directly damaged by the tsunami, together with Europe and the United States, proposed to set a

global alert system using meteorological satellites from the USA, Europe, Japan, China and India, and telecom satellites from India, Thailand and Malaysia. That was politically rather easy, generous and certainly not binding!

Then came August 2005! Just eight months after the Sumatra tsunami, the Katrina and Rita hurricanes probably contributed more than any political pressure to convincing the American people, although not easily, and hopefully the President and the rest of the world, that it was time to pay more attention to the vagaries of nature, and to take global warming seriously in hand even though, as discussed in Chapter 4, there is no direct simple connection between cyclones and global warming. In spite of its dramatic effects, Katrina and Rita demonstrated that even the richest nation of the planet might fall victim to one of the most deadly natural catastrophes with which it was ever confronted. The 1,800 victims of Katrina and Rita have probably not paid their heavy tribute in vain if they have been able to demonstrate that the climate may also change above the United States. Similarly, the droughts that affected California in 2007 and led to extended spectacular fires contributed to this *prise de conscience*. Things were also moving in other parts of the world. Major developing countries see climate change as a new barrier to their economic development. Argentina, for example, has suffered a 30% decline in its electricity production as decreasing rainfall and melting glaciers are reducing the reserves of its dams.

As of January 2007, in reaction to federal inaction, the Governors of eight US northern states – which altogether emit as much carbon dioxide as Germany alone – took part in the Regional Greenhouse Gas Initiative, a state level emissions capping and trading program [24]. In December, some 740 cities representing over 76 million people, including some large ones such as Baltimore, Boston, Dallas, Denver, Las Vegas, etc., started a nationwide effort to get other cities to support and agree to Kyoto and decided to jointly set targets for cutting back their own regional greenhouse gas emissions, even beyond the Protocol's targets. The state of Massachusetts is particularly sensitive to global warming as it might lose 300 km of its coastline as sea level rises. That surprising action was followed by the West Coast state governors, including Arnold Schwarzenegger of California (the 6th world economy and the 12th largest emitter in the world) who mandated a return to 1990 emission levels by 2020, and a further reduction to 80% of the 1990 levels by 2050!

The military also expressed their concern [25]. In March 2007, at the occasion of a colloquium organized to explore the strategic challenges created by global warming, John Ackerman of the Air Command and Staff College of the US Air Force, stated that war against terrorism should cede pace to sustainable security. Droughts, epidemic tropical diseases, water crisis, extreme climatic and weather phenomena, all potentially require military interventions as they might result in destabilizing humanitarian and social crises as well as massive migrations. The opening of the Arctic west passage, by creating a new maritime road, presents at the same time a new strategic challenge. The military response would result in more humanitarian operations, in adaptations of coastal infrastructure to sea-level rise and use of more efficient energy sources. As an officer requiring

anonymity said: '*Global warming is a reality, and the country and the army as well must prepare themselves*'! The sad example of the Darfur war, directly related to the effects of climate change, offered an unfortunate demonstration of what might be expected in other vulnerable parts of the world.

The business world itself seemed to move towards a more visionary approach to climate change. As it became more evident that the implementation of technologies contributing to reducing carbon emissions – such as better energy efficient buildings, hybrid cars, solar cells, wind turbines and nuclear power generators – would stabilize atmospheric CO_2 levels at their present value by 2054 [26], industry preferred taking the initiative now, investing in these cleaner technologies, rather than waiting for the government to force them into more expensive and painful processes. Not the least visible company in the United States, General Electric (GE) was pledging to double its investment in environmental research and development to some $1.5 billion a year by 2010 and to cut its own greenhouse gas emissions by 1% by 2012. Not much! But the belief of the GE Chief Executive was to invest *'in environmentally cleaner technology because we believe it will increase our revenue, our value and our profits, not because it's trendy or moral'*! Unfortunately, however, these examples were rather limited!

In retrospect 2007 may be seen as a turning point in the emotional perception of the issue. At the eve of the fourth IPCC meeting in Paris in January, came the undisputed news that the first six years of the century were the warmest ever recorded. The world public opinions became more emotional on climate change and started to think seriously about the measures they could adopt to contribute more efficiently to avoiding the disaster. For the first time, the voice of scientists could be heard loud and clear that the climate was warming beyond natural limits and that human activities were the cause of the phenomenon. The announcement in October that the Peace Nobel Prize was presented to both the IPCC and Al Gore put the issue on the front scene world wide and increased in an important way the emotional perception.

In the meantime, global warming became an election argument as the Bush administration was getting close to its term by the end of 2008. For the first time, the US Congress had started crafting comprehensive legislation to tackle climate change. If the two houses of the Congress were supporting some sort of action on global warming, the new administration could not ignore it. Furthermore, the new President could leave a positive mark in the history of his nation by adopting a courageous and visionary policy on climate change, being proactive *vis-à-vis* the European Union, the G8 countries and the major economies and developing nations, talking to them and proposing some concrete actions. In other words, regain the leadership!

Bali December 2007: a turning point?

The end of 2007 was crowned by the UNFCCC conference in Bali where some 10,000 delegates from 190 nations were supposed to reach an agreement on how to continue cutting CO_2 emissions after Kyoto in 2012. That Kyoto should have a successor was accepted by all parties: those who were in favor of continuing and

those who were against, and some were happy that it would be replaced by something more in line with their policies. The fact that the Protocol had not been able to reach its targets offered an argument to both the latter to fight for a change and to the former who wanted a re-enforcement of the limitations [17]. One may wonder why Montreal has been apparently able to curb the emissions of CFCs and initiate the recovery of the ozone layer, while Kyoto was less efficient. Was it in reality? Only one country was reluctant to admit its principles together with a few supporters sharing the same views of the future. In Bali, however, achieving unanimity and full consensus was considered as an absolute necessity especially when the reluctant partner is responsible for the largest share of CO_2 emissions. One element of the response is probably that Kyoto was dealing with the control of cheap energy consumption. As energy is both the source of subsistence and the motor of the most developed countries, it is not easy to accept that your food will be limited or that you will have to pay much more for it. In Bali, the majority was convinced that there was no way of going back to a pre-Kyoto era and that an international consensus on some kind of a treaty was to be achieved, with the developed countries leading the developing ones in order to implement a global strategy for curing the problem and curtailing CO_2 emissions. Bali was supposed to pave the way for an agreement to be reached in 2009 on a better Protocol or a more efficient approach than Kyoto. What happened?

Insults, threats, tears, booing, hissing and two sleepless nights at the end of two long weeks led to an agreement among all nations ... to talk more and lay a 'road map' for negotiating what would be the successor of Kyoto by 2009 allowing enough time to implement it at the end of 2012. Was this a 'historical' event? Certainly not by the substance of the agreement, but probably 'yes' because, for the first time in the history of the tortuous political discussions on climate change, the complete isolation of the United States led to a situation where the most resistant nation was forced to accept a final consensus. The world's largest CO_2 emitter was back at the table! For the first time in history, all nations, both developed and developing, the United States as well as China, came to an agreement recognizing that deep cuts in emissions were necessary to avert climate change, requiring all of them to do their share in limiting GHGs. Even though no specific numbers were agreed in the final report in order to satisfy the United States – contrary to what the representatives of the European Union were fighting for – and that no direct reference was made to the IPCC report (see reference [16]), just mentioned in a footnote, a cut between 25 and 40% by 2020 was mentioned if not prescribed. The action plan officially supported financial aid for efforts to prevent deforestation in developing countries. For the first time, it was recognized that the rich should help the poor to cope with the issue in particular by transferring climate-friendly technologies to them, even though the American representative refused a proposal for quantifying that technological assistance.

Bali could positively be considered as half-full rather than half-empty in spite of the apparent meager achievements. The feeling was that, after Bali, it would be

impossible to step back: progress was achieved! In Bali, the United Sates representatives implicitly admitted that the fight against climate change must be orchestrated within the United Nations framework. China agreed! The door was opened for real progress on the scale of the planet as the two countries responsible for the largest sources of climate deterioration understood that their own future was at stake. Climate change was already a debated item on the agenda of the forthcoming United States presidential elections. In the background, it was understood that it was wise to await the changes in Washington, as the forthcoming climate treaty discussed with the new administration in post would most probably rely on emission caps for industrialized nations. Definitely, the scene seemed to change!

11.4 Conclusion: towards world ecological governance?

Globality suddenly became a reality at the beginning of the 21st century. Artificial satellites have contributed in an important way to create that reality. The weather and climate affect nations well beyond their borders. It is noteworthy that the accelerated pace of climate change follows the most accelerated progress in technology research ever, and coincides with the advent of the space era. At the time when developed nations engaged in a non-controlled industrial expansion, they were developing the tools that would make it possible to control their development and render it sustainable, but it is only through science and education that this technological progress might uniquely contribute to the identification and the formulation of solutions to the present and future problems. The permanent monitoring of the state of the Earth will rest on an integrated system of satellites and associated ground-based systems – in other words, a world-scale GMES. The question is: To which organization should it report?

History tells us that crises have played a key role when it was necessary to create a consensus among nations – for example, the Society of Nations was formed after the First World War, and the United Nations after the Second. History also tells us that fear is a good motor for justifying and initiating long-term and global actions [27]. The control of the environment is increasingly perceived as urgent by many governments, with various degrees of seriousness however. As we advance into the 21st century the fear is visible as the problems become more and more real and acute. It is realized that it is time to think of a global management structure for saving the planet and ensuring a future to humanity as the Earth is the focus of the most severe attack by its almost 10 billion inhabitants, who are generously consuming its resources and dangerously affecting its climate.

Hopefully, some elements of the structure – of course, the easiest ones – are either already set up or slowly getting in place: weather and climate forecasting are dealt with at the planetary level through the WMO; space agencies are coordinating their programs and their projects through CEOS; and the GEOSS

has been created and is formulating its 10-Year Plan. The IPCC, in spite of its somewhat heavy structure, is nevertheless universally recognized as a necessity. Above all, its predictions can be confronted with reality, and are being heard more clearly and taken more seriously into consideration, as they prove the validity of the models on which they are based, showing the crucial importance of an in-depth scientific alert phase. A multitude of programs under the United Nations, whose list would be too boring to give here, are addressing the various problems that confront the planet and its nations. But when the political and the scientific worlds converge, and almost unanimously acknowledge that they share a problem, there is some room for optimism and hope about the century ahead!

In the wake of the IPCC meeting of January 2007, the outgoing French President, Jacques Chirac, invited representatives of 46 nations plus several Non-Governmental Organizations and representatives of the scientific and industrial world. They met in the Elysée Palace to discuss the concept of a United Nations Organization for the Environment which would replace the UNEP. The role of that new organization would be to evaluate and quantify the ecological hazards and their damages and to promote the measures necessary to safeguard the planet's environment. All the states of the European Union, plus several from Africa and Latin America, agreed to sign the President's call. But it was not signed by the United States or India or China. Obviously, some more time is necessary to achieve the proper perception! As said by Javier Solana, the Secretary General of the Council of the European Union: '*Global governance is an awful term but a vital concept. We need it because of a simple reality: interdependence*' [28]. A change is needed!

The question is whether this new global governance should report to the United Nations – as proposed by President Chirac – or be a newly created organization. Future governance must provide solutions at a global level but there is no consensus, in particular because of the USA's reluctance to delegate sensitive governance issues to the United Nations which they tend to relegate to a very subsidiary level, as illustrated by many examples. The United Nations Organization was established after the Second World War in a drastically different context, where globalization was not yet perceived as the challenge for the 50 years ahead and even less for the 21st century. At the beginning of the century, the increasing globality of problems is still in the hands of national policies and of the various governments; therefore, one has the right to question whether the present organization is adapted to the situation. Certainly, what the Earth does not need is to be trapped in a more bureaucratic system. Its problems require immediate and efficient actions, and very difficult choices and hard decisions have to be made!

The Earth needs an urgent recovery plan, together with an efficient management structure. The situation is not far from recalling that of Europe in 1945, devastated in some places upon all recognition. The President of the United States in January 1947 chose to appoint George C. Marshall as secretary of state who proposed a few months later his historical recovery program. In 1953, Marshall was awarded the Peace Nobel Prize in recognition of his contributions

to the economic rehabilitation of Europe after the Second World War and of his efforts to promote international peace and understanding. With the Marshall Plan, the United States in a most political and visionary manner re-enforced its role as Leader, saving Europe from further sinking, and accelerating its recovery. Today, in the context of globality, sovereignty and leadership are less and less manifested through the power of the army but through the ability of nations to sit around the international table – a table at which, occupying key positions, there should also be China, India and soon Africa, whose population is expected to reach more than 2 billion in 2050.

With the noticeable progress we have witnessed, we better realize that the hazards are not driven by fatality, but can be managed through education, scientific research, political dialogue and a willingness to act – and there is ample room for optimism! We are certainly far from having fully achieved that utopian goal, but we have witnessed a turn in the course of events, in the understanding of the problems and of their solutions, and in the recognition of the necessity to act now. In the next chapter we conclude this book, discussing the major issues and decisions that the planet and its populations are confronted with now, if we are to survive for a further 1,000 centuries.

11.5 Notes and references

[1] Baüer, P. et al., 2006, 'Observing the Earth: an international endeavor', Comptes Rendus Geoscience **338**, Elsevier SAS Publ. 949–957; and *Global Earth Observations Systems of Systems*, 2005, ESA SP-1284, p. 209.

[2] Committee on Earth Observation Satellites, 2005, *CEOS Earth Observation Handbook*, ESA, www.eohandbook.com, p. 212.

[3] ECMWF website: www.ecmwf.int

[4] The mission of the United Nations Environment Program, UNEP, is to provide leadership and encourage partnership in caring for the environment by inspiring, informing, and enabling nations and peoples to improve their quality of life without compromising that of future generations.

[5] GCOS is intended to be a long-term, user-driven operational system capable of providing the comprehensive observations required for:
- monitoring the climate system,;
- detecting and attributing climate change;
- assessing impacts of, and supporting adaptation to, climate variability and change;
- application to national economic development;
- research to improve understanding, modeling and prediction of the climate system.

GCOS addresses the total climate system including physical, chemical and biological properties, and atmospheric, oceanic, terrestrial, hydrologic, and cryosphere components.

[6] According to a report of the US National Research Council of 2007 *(Earth*

Science and Applications From Space: National Imperative for the Next Decade and Beyond), it is likely that, by 2010, the number of operating instruments on board NASA and NOAA Earth observation satellites will drop by 40%!

[7] Goetz, S., 2007, 'Crisis in Earth observation', *Science* **315,** 1767.

[8] CEOS (www.ceos.org) was created under the aegis of the Economic Summit of Industrialized Nations Working Group on Growth Technology and Employment.

[9] Bonnet, R.M. and Manno, V., 1994, *International Cooperation in Space. The Example of the European Space Agency*, Harvard University Press Publ., p. 163.

[10] 'Lower tropospheric temperature', *Science* **309**, 1548–1551.

[11] Nash, J. and Edge, P.R., 1989, 'Temperature changes in the stratosphere and lower mesosphere 1979–1988 inferred from TOVS radiance observations', *Advances in Space Research* **7,** 333–341.

[12] DMC images have been used to accurately measure opium cultivation in Afghanistan, which reached a record 165,000 hectares in 2006 compared with 104,000 in 2005.

[13] Richter, A. et al., 2005, 'Increase in tropospheric nitrogen dioxide over China observed from space', *Nature* **437**, 129–132.

[14] Nobuo Tanaka, Executive Director of the IEA, at the occasion of the launch of the 2007 edition of the *World Energy Outlook* in London on 7 November 2007, which focused on the energy developments in China and India and their implications for the world, said also: '*The huge energy challenges facing China and India are global energy challenges and call for a global response.'*

[15] *Environmental Effects of Ozone Depletion and its Interaction with Climate Change,* United Nations Environment Program, 2006 Assessment, and *Twenty Questions and Answers about the Ozone Layer,* a Panel Review Meeting for the 2002 ozone assessment led by W. Fahey, Les Diablerets, Switzerland, 24–28 June 2002.

[16] IPCC, 2005, IPCC/TEAP Special report*: Safeguarding the Ozone Layer and the Global Climate System: Issues Related to Hydro Fluorocarbons and Perfluorocarbons.* Summary for policy makers, Geneva.

[17] Prins, G. and Rayner, S., 2007, 'Time to ditch Kyoto', *Nature* **449**, 973–975. See also: the 'Green Climate Action task force' of the City of Takoma on the Kyoto Protocol at: http://www.cityof takoma.org

[18] The Exxon Mobil Company, secretly reported to support several groups that were seeking to cast doubts on the science of climate-change, issued the opinion '*that the scientific evidence on greenhouse-gas emissions remains inconclusive and that studies must continue while tangible actions are taken to address potential impacts'*. Between 1998 and 2005, Exxon Mobil has been charged by 'The Union of Concerned Scientists' (an independent science-based non-profit group working at securing the environment and safety, combining independent scientific research and citizen action to develop innovative, practical solutions and to secure responsible changes in government policy, corporate practices, and consumer choices) to have distributed some US$ 16 million to bodies dedicated to amplifying public

perceptions of the scientific uncertainties over climate change. See also: 'Exxon Mobile accused over strategy on climate change', 2007, *Nature* **445**, 137.

[19] Dennis, C., 2006, 'Promises to clean up industry fail to convince'**,** *Nature* **439,** 253.

[20] Stern Report, 2007, *The Stern Review on the Economics of Climate Change,* Cambridge Univ. Press, p. 712. The report stated that acting now would cost much less than later. It predicted that between 5% and 20% could be wiped off the global Gross Domestic Product (GDP) by the beginning of next century if nothing was done, while the costs of reducing could be limited to around 1% of the global GDP each year. It proposed a global coordination, insisting on the necessity to double the research budgets in the field and to involve not only the rich nations but also the developing economies.

[21] GMES, http://www.gmes.info/157.0.html

[22] Butler, D., 2007, 'The planetary panopticon'*, Nature* **450**, 778–781.

[23] Lubick, N., 2005, 'Something to watch over us'*, Nature* **436**, 168–169.

[24] The Regional Greenhouse Gas Initiative, or RGGI (http://www.rggi.org/), is a cooperative effort by Northeastern and Mid-Atlantic states to reduce carbon dioxide emissions. In that perspective, the RGGI participating states intend to develop a regional strategy for controlling emissions that will more effectively control greenhouse gases, which are not bound by state or national borders. Central to this initiative is the implementation of a multi-state cap-and-trade program with a market-based emissions trading system. Currently, seven states, including Connecticut, Delaware, Maine, New Hampshire, New Jersey, New York, and Vermont, are participating in the RGGI effort. Legislation was signed in April 2006 requiring Maryland to become a full participant in the process by 30 June, 2007. In addition, the District of Columbia, Massachusetts, Pennsylvania, Rhode Island, the Eastern Canadian Provinces, and New Brunswick are observers in the process.

[25] Busby, J.W., 2007, *Climate Change and National Security. An Agenda for Action*, Council Special Report No. **32**, ISBN 978–087609–413–6, Council on Foreign Relations, p. 40.

[26] Pacala, S. and Socolow, R., 2004, 'Stabilization wedges: solving the climate problem for the next 50 years with current technologies', *Science* **305**, 968–972.

[27] According to Jacques Attali, former adviser to the French President François Mitterrand, the setting up of the European construction – starting in 1947–1948 with the cold war and ending with the fall of the Berlin Wall in 1989 – was built on four main fears: the return of Nazism in Germany, return to the shameful French Government cowardice during the War, the Soviets, and the departure of the US army. After these fears had disappeared, the European construction faced more difficult problems 50 years after the war than just after the war! (*Le Monde*, 7–8 January 2007.)

[28] Solana, J., 2007, 'Countering globalization's dark side', *Europe's World* **7**, 114–121. Available on line at http://www.europesworld.org

12 Conclusion

Our hopes for the future state of the human species can be reduced to three important points: the destruction of the inequality between the nations; the progress of the equality within one people; finally the real perfection of mankind . . . where the stupidity and the misery will be only accidental and not the usual state of a part of society.

Marquis de Condorcet

For many millions of years life on Earth has carried on through changes of atmospheric conditions, warm periods and glacial epochs. At times major catastrophic events occurred such as volcanic eruptions or meteoritic impacts which killed off significant parts of the biological world. But the potential for a further flowering was always preserved. Of course, we humans are not just concerned with the continuity of life in general, but more with the continuity of our own species. The great apes and their evermore sophisticated descendants have existed for quite a few millions of years and anatomically modern humans for 100–200 thousand years. So, unless a universal nuclear conflagration destroys us, or if an end-of-Cretaceous type impact strikes unexpectedly, there seem to be no grounds for believing that our species could not last another 100,000 years. But this does not say how many individuals of our species the Earth can support for a long period.

12.1 Limiting population growth

In many animal species the number of individuals is regulated by the availability of food and by the presence of predators. For humans the latter play a negligible role. Without fertility control they therefore would multiply until starvation limits the population. While there have been many opinions as to the number of people that can live on Earth, if population continues to grow, overcrowding or starvation would result very early in the 100,000-year period. So it is abundantly clear that a long-term stable population is an absolute necessity for humanity to survive in 'decent' conditions and to preserve human dignity. As discussed in the Introduction, the medium United Nations forecast is for the population of the Earth to stabilize at some 11 billion people. In subsequent chapters we have seen that probably our agriculture can feed that many people and that the necessary natural resources from water, to energy and iron are obtainable. But we have also

seen that this will not be easy and will depend on a highly efficient, environmentally friendly, agriculture maintaining biodiversity and on an ample availability of energy. Certainly the problems would be a lot easier to handle with a smaller population.

The richer countries, such as Japan, have managed to delay their population increases while the poorer countries have not. There are good reasons for that, the most logical being the need for poor countries' families to count on a larger workforce per family to secure their needs; another being the lower level of education. Some countries, such as China, have adopted a more authoritative approach. On the other hand, most of the European countries, and some others, are witnessing the effect of education and individual openness to birth control. The benefits are visible: if, in China, the population had continued to grow as it did in Africa from 1965 to 1970, there would now be 2,100 million Chinese, 800 million more than the actual number. Seeing the actual problems of the country we could imagine what the consequences would have been. But now the negative aspects are beginning to show up: aging populations and stress on pension and health systems. In some countries there is even political pressure to increase the birth rate through special rewards. As the population cannot continue to increase, this appears to be shortsighted. In the past 100,000 years, wars, epidemics and natural disasters have been unable to limit the population: the grand total of all fatalities attributed to these disasters has reached at most a few hundred million over the last century. This is to be compared to the increase of 4.4 billion in the world's population in the same period. If humanity is to continue to exist under tolerable conditions, the sociopolitical system has to be adapted to a society without further population growth. And the sooner we ensure this the better for all, if one does not want to have a world without forests, beaches, lakes and rivers – that is, a world without 'Nature' – as well as the evident problems with agriculture and pollution. If we were to grow for one more century from 11 billion with a 1.5% per year population increase, we would arrive at 30 billion people living on half a hectare of the Earth's surface per capita. What we will do in the coming 50 or 100 years will determine the living space offered to us in the long-term future. In this respect the effects of increasing population and of continued CO_2 emissions are very similar: they do not just affect the countries that cause the harm, but the whole world. Neither of them respects national borders. Already in the present world one sees that it is impossible to confine people to areas of poverty and overpopulation. They will break out, and unless one constructs a brutal world with walls everywhere, containment will not work. Some of the world's great religions have been indifferent to the essential importance of limiting the Earth's population, or have even actively opposed it. Can they continue to play with the fate of the Earth without concern for the well-being of their own adherents and others? And is it really more in accordance with the precepts of religion to shoot those who try to escape from misery than to limit the population? It is one or the other.

Whether we like it or not, the Earth is the planet of the future for the whole of humanity, which has no other choice but to live on it and feed from it, and

therefore has the duty and responsibility of preserving it as its only possible habitat for the next 1,000 centuries. But this is not granted, and requires some drastic changes in the way we humans have developed on Earth until now, since we became 'modern'. This lack of alternative to the future will undoubtedly raise criticisms from the believers in the possibility of ensuring our future on other planets or in space habitats. 'Extraterrestrialization' has been invoked as a means of regulating the Earth population in the space age. The concepts of colonizing Mars and other worlds, as well as the building of artificial non-planetary habitats, are still in the realm of utopia. They are unable to solve the problems and the future of our civilization. Space colonization will not be able to offer a global alternative to a dying planet because of the costs, the risks and the physical limitations. Even though we might for a moment assume that, at some time in the future, when we possess the technological ability to do so, it is decided to terraform Mars and live on it, the Earth will not remain without humans as the population of the Earth will be in no worse position than the new Martian population. If we could attain that level of technological development, we would also have all the means to control the demographic, technical and industrial development on Earth and maintain it in a liveable state.

This does not imply that we won't live on Mars or on other objects of the Solar System, but we will do it 'attached' to the Earth and not independently or autonomously. We need the mother planet as a place to which we can continue to return after our sojourns – assuming, first, that we survive the more dangerous and artificial environments of these mythical extraterrestrial habitats and, second, that the Earth itself is maintained in a liveable state.

No one should believe that, in the present strongly unbalanced situation of terrestrial populations, with the rich making money and the poor making children, we could solve the problem of survival of all by 'extraterrestrialization'! Those who tend to promote such concepts seem to refuse to face the reality of an immediate predicted disastrous future that requires drastic redirection of the global management of the Earth. 'Extraterrestrialization' is in no way a solution to our future problems. It might be considered in a situation where all the problems have been solved. The Earth deserves more attention than the Moon presently! One is dead; the other should survive! We therefore repeat once more that there is no alternative to limiting the population. We admit that this is not easy since no presently acceptable method can be imposed to achieve such a goal.

12.2 Stabilizing global warming

A rapidly increasing global warmth and rising sea levels will create problems for large areas of the globe, which will be more difficult to solve, the larger the numbers of people involved. Dangerous climate change is likely within a century if we do not control our consumption of hydrocarbons. Again, this is an irreversible process; even if later we stop using them, the temperature would continue to increase still further, the ice on Greenland and west Antarctica

would start on a melting trajectory and climate might be locked into a different warmer state. In the case of an event 55 million years ago we have seen that it may take the better part of 100,000 years to revert to anterior conditions. Similarly, if we continue to cut the tropical rainforests in Amazonia or Indonesia, a permanent change with greater dryness may well result. So many examples could be given that the present century may be critical for the well-being of humanity for many years to come. Since power plants have lifetimes in the order of 30–50 years, we really do not have much time left to change direction away from coal and other hydrocarbons.

12.3 The limits of vessel-Earth

We admitted at the beginning of the book that forecasting the future may well appear as somewhat irresponsible to the reader. Such an endeavor can be based only on models, which is the approach taken by the IPCC to assess the evolution of the climate and the effects of global warming. Analyzing the conditions for our survival is of a similar nature, be it on Earth or even on Mars or on any of the possible bodies of the Solar System: together, this set of habitats constitutes a closed system with limited resources and, mathematically, the consumption of non-renewable resources will tend towards zero with time. We may also use a very simple, and often chosen, model to explain the situation. This model assimilates the Earth to a space vessel orbiting autonomously in the Solar System. Because the usable resources are limited, the vessel cannot expand, the number of astronauts cannot increase, inequality between them is incompatible with the goal of surviving the journey, while its management or governance, by essence, has to be rigorous and decisional as it must keep the system safe and hospitable for all its inhabitants, otherwise, the 'vessel' will be on the dangerous track of self-destruction. Our future requires that 'vessel-Earth' is maintained in a situation where the limitations are neither disputable nor negotiable, and non-acceptance of this fact is equivalent to adopting collective suicide as the ultimate fate of humanity. We do not accept such a fatalistic view. Therefore, we have no choice other than to maintain 'vessel-Earth' in a liveable state, and accept that this vessel is the only one able to accommodate a crew of several billion 'terranauts' in the most biodiverse environment possible – the only one we have.

At the end of our exercise, not surprisingly, we find ourselves in full agreement with the conclusions of the study of the 'Club de Rome' and of their *Limits to Growth*. The finite resources of 'vessel-Earth' and the existence of global warming impose, more clearly now than in 1972 when the study was published, a new approach to our economic systems and to the global management and usage of these resources: a perpetual material growth will lead sooner or later to a global collapse. How soon? How late? The System Dynamic Model adopted by the Club de Rome, and used by the Massachusetts Institute of Technology to establish the conclusions of the report, probably came too early as it could not benefit from the fantastic development of informatics witnessed in the last 20 years of last

century. The exercise might be worth revisiting on the basis of the evolved situation since 1972. This may tell us more accurately how soon the situation will become unliveable for different scenarios, but it will not change the laws of mathematics or physics and therefore not the conclusion that our way of doing things must follow a different track. And the situation is worse than it was when the exercise was first conducted. Today global warming exists. The evidence for limiting the growth of at least the consumption of fossil fuels and of the released greenhouse gas emissions is imposing itself every day more strongly to the world, even though the approaches on how to achieve these limitations differ deeply from country to country. At the end, however, all nations must share the responsibility of ensuring the success of these limitations and eventually the benefits of a world where disparity will fade away, with better chances of balance and hopefully peace.

This book also shows that in many areas we are hitting the limited volume of the 'vessel' and the means that exist on board to secure its future and that of its crew. It is becoming worse, but is recoverable, provided that the proper measures are implemented soon in terms of usage limitations: the recycling of natural resources of water and minerals, and the preservation of soils and renewable energies. Until now, the natural tendency of governments has been to trust that our technological abilities will offer unlimited, but undefined, opportunities to progress. Quoting *Limits to Growth,* no one, however, can dispute the impossibility that '*material growth can go on for ever*'. The urgency of adopting the proper measures is appearing every day more striking than yesterday. In particular, the inertia of the 'system', both natural and political, tells us that even if we undertake to correct the present course of action now, the effects of these corrective measures will not be evident for several decades. This is the case of population growth, of greenhouse gas emissions continuing their dangerous effects on global warming, and of the increase in the ozone-hole expansion. Furthermore, all our data indicate that the longer we wait, the more difficult it will be to adopt later solutions that would lead to stabilization and, as discussed explicitly in the Stern Report on global warming, the more costly to the world GDP will these measures be.

12.4 The crucial role of education and science

The activities and the role of the IPCC have evidenced that the political world needs the best brains to constantly and properly assess the state of the Earth. Research and technological development are required in all actions connected to environmental activities and space systems, as technological advances are necessary to reach the world equilibrium state – a world that needs constant monitoring from space and most probably a more innovative and continuous effort on geo-engineering. For the next 1,000 centuries civilization will be dominated by scientific knowledge and by technologies in order to be able to improve our abilities to model and forecast our future and develop more efficient

means of saving resources or, possibly, tapping into other sources as necessary to ensure survival.

A World Global Monitoring Environmental Agency (WGMA) seems to be a possible model that would fund and orchestrate the development of all indispensable means, including the space means, to properly manage the running of the 'vessel' whose permanent evolution and complexity requires permanent evaluation. This agency would be financed by all nations and would ensure the availability of the best scientific expertise in the world. It should also be in a position to identify the necessary technologies and invest in areas concerning energy sources, water management, agriculture and transport. It should secure the permanence and uninterrupted presence of monitoring and observation tools, be they located at the surface of the planet or in space. It should have the capability of politically-independent assessment of the evolution of the physical situation of the planet where the evidence of certainties, or their lack, should be established and accepted as normal and not as opportunities for no action. It should offer a continuous assessment of the validity of these pre-visions through a comparison between their anticipated results and the real and measured values of critical parameters for the period considered. There exist in Europe a few elements that might constitute the principles and the basis of such an agency: the European Space Agency, Eumetsat, and the European Center for Medium Range Weather Forecasts. All are involved in the definition and development of the Global Management Environment and Security (GMES) program – a possible seed for a similar program involving all nations of the world.

It is all the more upsetting that, at the time when research is so critical, fewer and fewer students in schools and universities are considering scientific careers while, at the same time, unambiguously recognizing the importance of science and technology for society. This illustrates the necessity to re-valorize the scientific careers in all countries, so that they become more attractive to the future generations. Hope in that direction has come from the recent 'Rose Report', published recently under the aegis of the University of Oslo [1]. This report shows that, as far as education is concerned, the poorest countries may have an enormous lag but they realize the importance of education – and of science in particular – to secure their future and offer them more optimistic prospects. In that perspective, education will make it easier for these countries to adopt a less pressing attitude towards increasing the number of children and thus avoiding going through a dictated authoritative approach.

12.5 New governance required

As we have mentioned several times in this book, global problems can be solved only on a global basis. The founding of the United Nations after the Second World War, and more specifically in our areas of concern, the Montreal and Kyoto Protocols provide examples of the growing awareness of the need for

global solutions. The implementation of the required measures to address these problems properly calls on a new governance of the world in all domains where the global future is at stake. The difficulty will be for the most powerful and richest countries to admit that they have to limit their level of living to enable the less rich countries to reach a more equivalent level of life. However, trying to implement that too soon may not lead to the expected result. Inventing new institutions to govern the future world is a risky game and the problems encountered in building Europe – a small model for the world? – evidence the difficulties of such a model when it is considered to extend it on a more global scale. The most dangerous trap is to invent new bureaucracies whose aims might just to offer seats to some ineffective government employees.

Nations and governments more than often face extreme difficulties in thinking a few years ahead and above the level of their peculiar immediate interests. Imagine centuries or thousands of centuries! In any event, something must be done and the near imminence of a deep crisis may force these nations to act responsibly, as we have learned from the past that only the pressure of crisis forces them to find the right compromises and solutions to their problems. The attitude of the man jumping from the top of the Eiffel tower saying 'so far so good' when he reaches the second floor, does not lead to any issue other than an ultimate and fatal crash! The race to growth presently involving nearly all countries will lead to an unavoidable impasse at some time during this century or the next. As shown by the 'GEO 4' report of the UNEP issued at the end of 2007, the future of our environment in the next 50 years is directly determined by the socioeconomic model that will be adopted. The report, prepared by several international experts, clearly outlines that ecologic and sustainable development – be it for climate, energy, water, biodiversity and equality – is incompatible with the model of an unlimited growth. Whether we like it or not, the era of wilderness that governed the running of the world until now will be over sooner or later, and most likely in this century. Trying, however, to solve all problems at once may be counterproductive and a more step-by-step approach is probably more easily acceptable. Nevertheless, actions are necessary, and the sooner the better.

We are, however, comforted in this somewhat idealistic conclusion by what we witness on the front of global change with an increasing number of countries willing to address the issue responsibly – with Europe clearly taking courageous and advanced positions – envisaging the development of some of the tools we discussed above such as the United Nations Environment Organization recently proposed by the French President. The World Bank, the International Monetary Fund (IMF), and the World Commerce Organization, are re-evaluating their policies towards less liberalism and more responsibility of governments in the areas of agriculture, infrastructure and social protection in view of transforming growth into sustainable development. Industry is also exploring the promises of new approaches to growth. On smaller sociological scales, we also witness initiatives for achieving well-being in a world not dominated by growth and consumption, like Economy Nobel Prize winner K. Arrow asking whether we are

not consuming too much (*Journal of Economic Perspectives*, 2004)! Starting also at the individual or family levels, new ways of living, new approaches to management of needs and resources, as well as of time, are developing which may make it easier in the future to broaden the approach at the level of cities, nations and then the whole planet.

The most obvious organization to establish the new governance mentioned earlier, the proposed WGMA, would be the United Nations. On the political level it was recently suggested that the G8 and the Security Council should be merged [2]. Referring to environment issues, we also quoted the WMO and the IPCC as good examples of institutions and programs that are presently the best available, although insufficient, approximations to what we think will be necessary to manage the world in the centuries to come. The one important problem, however, is that not all nations are ready to accept a common discipline or to adopt a consensually approved behavior. That is nevertheless necessary. Piloting a space vessel requires not only the acceptance of a consensus on the part of all crew members but also some leadership responsibility. The allegory is valid for the management of 'vessel-Earth'. The vessel must be driven properly, with all countries agreeing on both the state of the vessel and on the solutions necessary to ensure its ability to host its populations. This leads us to think that some adaptation of the UN system of agencies and programs to the challenges we are facing – where the security might shift at some stage from the military agenda to the civilian agenda – is probably the less difficult solution.

12.6 The difficult and urgent transition phase

We have seen that it is possible to imagine a pleasant world that can exist for a long time with adequate water, food, energy and other resources. It is important to realize that this possibility is contingent upon measures being taken during the present century or even during the coming decades. This is obvious for the necessity to limit the world's population. We have seen that in water usage and in food availability the 11 billion people projected for the Earth are straining the possibilities. So it is essential to induce the countries that still have rapidly expanding populations to take the necessary steps as a high priority; and it is equally essential for the survival of the tropical forests and their denizens. The climate and the sea level would be irreversibly changed if measures to limit CO_2 production were not taken now, so it is imperative to have a more constraining version of the Kyoto Protocol adopted very soon. But what makes the transition phase so difficult is that the different aspects are all interrelated. To have adequate food, it is necessary to have adequate land and water. Long term, with better crops and with adequate water, it will be possible to grow adequate amounts of cereals on the present agricultural land, and so prevent deforestation. But today we do not have these crops and we do not have the energy to desalinate sufficient water from the oceans. Unfortunately, it is today that more food is needed because of population increases, and people will therefore cut

down the forests to gain more land for agriculture, even if that damages the long-term productivity of the soil. But even if that land was no longer needed it would not restore the forest. In this particular example there is a possible solution for the immediate future – namely, to utilize more fertilizers – but the countries that require the food have no money for fertilizers. If we consider that it is in the interest of the future world to save the forests, the richer countries will have to provide the fertilizer for free. More generally, it will be necessary to transfer substantial resources to the poorer countries and speed up development to the maximum, if only to ensure that population growth is moderated rapidly. Otherwise, we may reach the point of no return where the population is too large to make a benign future possible.

12.7 Adapting to a static society

Humans have not only physical needs. Can their minds survive the conditions of a 100,000-year world? We have found that in such a world any important change is difficult. Even a 1% per year growth is impossible. So we have to adjust to a more static society. During human history two societies have had such a character for several millennia – the Egyptians and the Chinese. Both were very unequal societies, and in that respect very different from what we have assumed. In Egypt a small class of nobles lived well during the pleasure of the pharaoh. Unemployment of the numerous proletariat was prevented by putting them to work on 'pharaonic' public works, thereby not leaving much leisure for developing ideas for change. In China, in addition to the hereditary nobility, a class of functionaries was selected by merit. It is reported that to fill the post of chief adviser to the emperor some 400,000 candidates were vetted first at the provincial level and finally in a national competition. Since the subjects of the competitive exams consisted mainly of a knowledge of ancient texts, conservatism was largely guaranteed. In fact, some deliberate measures were also taken – such as the dissolution of the fleet of Cheng Ho that had made seven grand voyages with some 60 large ships (1405–1433) – perhaps to ensure that new ideas would not enter and endanger stability.

Somewhat less long lived, the Roman Empire was far less stable in part by a geography that gave it a large periphery of hostile, subjected areas. For a rather brief period it showed another way of dealing with the proletariat: providing them with free food and amusement, the productivity having been displaced to the subject populations in the periphery. The situation was inherently unstable since much of the military power had also been moved to the periphery.

For the last 500 years European society has been operating in a very different mode: growth and change involving scientific and technological development. It has given humans an unparalleled power over the natural environment, and gave the technologically developed societies a fair power, but far from absolute, over the less developed ones. Two consequences are in evidence: a quasi-absolute belief in continuing growth and progress (1) without an idea of what the end

result should be or (2) with an idea that there is no final point. Also the belief that, irrespective of the problems that may be encountered along the way, technological fixes will be found for their solution. All of this was further fostered by the fact that the finiteness of the area of the Earth was not a problem for the Europeans. When there was no longer space at home, huge areas could be colonized elsewhere: North America, Australia, Africa and South America. The Egyptians and the Chinese had a different experience. Their territory was finite and at least the latter preferred it to remain so. The European experience was unique in the history of the world and cannot be repeated, since 'empty' lands with favorable conditions no longer exist.

In the European and, even more, in the American mind set there remains a brilliant future on the horizon, with great vistas of scientific discovery and technological development as well as economic growth. Characteristic for postwar science was the title of the report *Science, The Endless Frontier* [3]. But will scientific development really be 'endless'? Of course, we can continue to build ever larger particle accelerators or telescopes, and undoubtedly we will continue to do so for some time. But will the discoveries we make have the same impact on our image of the world as those of an earlier generation? We doubt it. The development of quantum physics was so exciting because it fundamentally changed not just a specialized part of physics but all of it, as well as chemistry, astronomy and biology. Nuclear and particle physics also affected science and technology in a fundamental way. And the discovery of the expansion of the Universe, its immense size and finite age, changed our view of the place of humanity in the scheme of things. The next steps in particle physics and cosmology are unlikely to have a similar impact, because these do not affect other sciences to any great extent. We may not be so affirmative in the fields of biology, genetics and cognitive sciences which open new avenues, with the potential of many subsidiaries leading to new yet unidentified problems, of which the Genetically Modified Organisms offer an example.

Another one comes from the social behavior of future civilizations in a world that is forced to adopt the philosophy of the limits to growth and is somewhat constrained by common rules which, at least in appearance, leave less room for individualism and complete freedom of action. In other words: what shall we do in the next thousand centuries? This question was also raised in *Limits to Growth* and it is to be noted that these problems have become more openly discussed in publications and large audience newspapers. These limits do not concern any of the human activities that are not dependent on the usage of non-renewable resources or on anything that would severely damage the environment. We can list just a few obvious things that come immediately to mind: scientific and technological research, space exploration, discovering and imaging new worlds, education, art through all its forms, sports competition, space tourism and search for new resources, politics, not forgetting religion in as much as it rests on mutual tolerance. In other words, societies should invent a new equilibrium in rebalancing active life in the direction of more creation, politics and contemplation. We believe, perhaps too optimistically, in the power of the

human brain above that of the animals, and in its inventiveness in developing tools that have reached, as we presently witness, a level of power and sophistication unthinkable 100 years ago for finding solutions to our fate and that of the future generations. These solutions may appear naturally and continually as we learn how to live within new constraints, and benefit from a new equilibrium based on new and global ethical rules, as part of new forms of society. For that, we must be convinced that this is the only long-term goal and that we have the will to achieve that goal. This will be somewhat difficult to imagine and to accept because we have never experienced such a drastic change.

The 21st century is a unique test case for our ability to survive longer. It is the century of globalization and the first to fully confront humanity with the negative effects of blindly driving the world – a world on which we are rapidly reaching all limits to growth. It is a narrow path in time that all nations must successfully go through collectively and not individually. If they manage that difficult transition they may have a better chance of surviving the subsequent centuries and, perhaps, for 100,000 years more. This test case should be successful in all circumstances: there is no other choice. To the question 'Can we survive a thousand centuries?' we must admit honestly that we cannot today offer a positive answer because that answer is not only ours but must be given collectively. We do believe, however, that before the end of this century, humanity will be in a position to propose an answer that will, hopefully, be positive.

12.8 Notes and references

[1] Sjøberg, S. and Schreiner, C., 2007, Rose Report, University of Oslo.
[2] Attali, J., 2007, *Le Monde*, 7–8 January.
[3] *Science: The Endless Frontier*, 1945. A Report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development, July 1945, United States Government Printing Office, Washington.

Index

ADM–Aeolus satellite, 370
Aerosols, 109–111, 191, 324, 326, 329, 352, 354, 359, 370, 374
Africa 262, 266, 270, 271, 274, 404
African monsoon, 136, 144
Agassiz, Louis, 153
Agriculture, 7, 223, 263, 324
AIDS, 97, 98, 147
Aitken basin, 19, 62, 63
Albedo, 64, 69, 283–5, 290, 320, 324, 326
Alberta 233
Algae, 37, 225
Algal blooms, 264
ALOS satellite, 345, 348
Altimetry, 319, 322, 331–8
Aluminum, 243, 248
Alzheimer disease, 98, 101
Amazon, 144, 255, 256, 332, 333
Amazonia, 144, 270, 272, 274, 275, 333, 336, 406
Antarctic, 161
  Ice Sheet, 161, 199–201
  Ice shelves, 157, 171, 199–201
  Peninsula, 153, 200
Antimony, 244, 248
Apollo, 19, 21, 22, 282, 299–301, 304
Apophis, 78, 79, 83, 84
Aral Sea, 144, 259
Archea, 37
Archean, 35
Arctic, 171, 172, 198, 272, 351
Arctic sea ice, 199
Argo system, 322
Arrhenius, Svante, 160
Arsenic poisoning, 260
Association of Southeast Asian Nations (ASEAN), 388
Assuan dam, 258
Asteroids/comets, 62–85, 241, 281
Aurora borealis, 125, 176
Bali meeting, 395–7
Bam earthquake, 343, 345
Banded iron formations, 37
Bangladesh, 134, 143, 207, 259, 322, 341
Baptistina, 45, 46
Basalts, 244
Beppi Colombo mission, 76
Bering land bridge, 154
Biodiversity, 272, 298, 316, 327, 328, 368, 376, 384, 391
Biofuels, 213, 223–5
Biomass, 217, 235
Biosphere, 316, 320, 324, 327–9, 329
Biosphere project, 299
Black Sea, 126, 128, 138
Boreholes, 172, 173
Borneo, 274
Brazil, 386
Brazilian environmental laws, 274
Brundtland Report, 3, 384

C3/C4 plants, 269
Calcium-Aluminum-rich Inclusions (CAI), 16
CALIPSO mission, 354
Cambrian, 17, 35, 42
Cambrian explosion, 17, 41
CAMP, 46, 47
Campi Flegrei, 114
Carbon-13, 36
Carbon-14, 15, 176
Carson, Rachel, 264
Caspian Sea, 258
Catarina, 133
Cenozoic, 18
Central America, 195, 262
Centrifuges, 227
Cereal production, 267
Cereal yields, 266, 267, 269
Chain reactions, 226

Challenger accident, 299
CHAMP mission, 332
Chernobyl accident, 225
Chi-Chi earthquake, 124
Chicken, 266
Chicxulub, 44, 62, 63, 65, 67
Chile, 118, 119, 126, 268
China, 5, 116, 119, 122, 125, 126, 134, 140, 142, 143, 257, 262, 266–9, 274, 353, 354, 361, 378, 379, 382, 383, 386–8, 393, 394, 396–9, 404, 411
Chlorofluorocarbons, 33, 111, 180, 182, 292, 325, 382, 384
Cholera, 347
Chondrites, 32, 241
Clementine, mission, 19
Climate 6, 156–9
Climate change in Amazonia, 275, 276
Climate forecasting, 189–96, 201–6, 373
Climate Models, 189, 373
Club of Rome 4, 406
CO_2, anthropogenic, 201–4, 208, 233–7, 263
  natural, 42, 46, 158–67, 177–9, 187–94
  on planets, 282–92
  sequestration, 236, 308, 309
Coal, 232–236, 382
Cobalt crisis, 243
Committee on Earth Observation Satellites (CEOS), 377, 378, 397
Concorde aircraft, 373
Copper, 243–245, 248, 268
Coronal Mass Ejections (CMEs), 359, 361
COSMIC mission, 339, 380
Cosmic rays, 53, 54, 57–61, 89, 176, 317, 359, 361
Craters, 13, 19, 22, 62–66, 70, 72
Cretaceous-Tertiary (K-T) extinction, see K-T
Crutzen, Paul, 309
Cryosat, 336, 340
Cryosphere, 320, 321
Cumbre Vieja, 131
Cyanobacteria, 35, 37
Cyclones/hurricanes, 94, 132–7, 140, 147

Dansgaard-Oeschger events, 168
Darwin, C.G., 1
De Orellana, Francisco 272
Decarbonization, 214
Deccan traps, 46, 105
Deep Impact mission, 72
Deforestation, 224, 264, 273
DEMETER, 124, 361
Dengue fever, 347
Desalination, 6, 246, 260–2
Desert, 265
Deuterium, 226, 228–32, 285
  δD, 163, 164
DIA, 254, 257, 262
Digital Elevation Models, 343, 344
Dinosaurs, 43–7, 62, 74, 75
Disaster Monitoring Constellation, 379
Distribution of warming, 172, 194–196
DNA, 100, 101, 297
Dobson Units, 180, 351
Domestication, 263
Don Quijote, mission, 82
Droughts, 143–6, 195
Dyson, Freeman, 286, 308

Earth, 20, 21
  atmosphere, 31–4
  core, 16, 17, 20, 31
  dynamo, 28, 317
  energy budget, 355
  history, 41, 286
  Impact Database, 63
  magnetic field, 28–31
  mantle, 20, 21
Earth Observation Summit, 392
Earthquakes, 25, 93, 94, 102, 112, 115–28, 130, 138, 147, 318, 319, 338, 343, 345, 347, 369, 372,
Eccentricity, 166, 167, 292
Ediacarans, 38–43
Eemian, 170, 201
Effective temperature, 160, 283, 284
Egypt, 139, 411, 412
El Niño, 136, 141, 144, 162, 371, 373
Elements, abundances, 238–49
Energy, cost of, 220, 221, 225, 230, 237
  renewable, 6, 235, 236
  reserves of, 234
ENVISAT, mission, 328, 334, 336, 341–55
EPICA, 163, 165, 170
ERS, missions 334–7, 342, 344, 346, 351, 383
Ethanol, 223, 224

Etna, 104, 343, 346
Eumetsat, 369, 370, 372, 390
Europa, 294, 295
European Center for Medium range Weather Forecast (ECMWF), 370–2
Extinctions, 17, 18, 43–8, 53, 146

Faint young Sun, 33–4, 160, 285
Fast Breeder Reactor, 227, 228
Feedback, 158
Fermi Paradox, 299, 300
Fertilizers, 223, 248, 267, 270, 411
Final energy consumption, 216
Fission, 226
Flood basalts, 46, 47
Flood waters, 255, 256
Floods, 137–42
Food production, 7, 265–71, 276, 410
Forcing, 158, 190, 191
Forest fires, 44, 45, 144, 275
Forests, 265, 271–6
Fusion energy, 217, 228–31, 304, 305

G8 summits, 388, 392, 393, 395, 410
GAIA astrometry mission, 76
Galileo satellites, 337, 338, 390
Gamma ray bursts, 60, 61, 357
Ganges, 122, 139, 143, 157, 207, 263, 341
Gas, see natural gas
Genetically modified organisms, 264, 269, 412
Geo-engineering, 308–10
Geoid, 331, 332, 334
GEOSAT, 334, 335
Geostationary orbit, 302, 305, 306, 330
Geothermal energy, 214, 217, 218, 235
Gibraltar Strait, 138, 143
Giotto, mission, 70–2, 80
Glaciers, 189, 207, 263
Gliese-581, 296, 297
Global Observing System (GOS), 378, 379
Global Earth Observation System of Systems (GEOSS), 380, 391–3, 397
Global Monitoring Environment and Security (GMES), 388, 390, 397, 408
Global Positioning Systems (GPS), 319, 332, 334, 337, 338, 359
Global Precipitation Measurement mission (GPM), 380
Global warming, 171, 172, 192, 193, 392, 393, 405–7
GLONASS, 337
GOCE, 332, 333
Gold, 242, 244, 245, 248
Gondwana, 25, 27, 161, 234
GOSAT mission, 324, 353
Governance, 368, 384, 397, 398, 408, 410
GRACE satellites, 198, 320, 331–4
Granites, 23
Gravimetry, 320, 331, 332
Great Oxydation Event, 39
Greenhouse effect, 33, 34, 57, 160, 282, 284, 285, 287–92, 294, 308, 324
Greenhouse gases, 33, 45, 75, 108, 110, 160, 178, 188, 189, 292, 293, 308, 309, 320, 324, 326, 327, 352–5, 373, 374, 382, 384, 388, 394, 407
Greenland Ice Sheet, 161, 170, 187, 197–9, 201, 204, 205
Greenland settlements, 154
Ground water, 254, 258, 260
Group on Earth Observation (GEO), 391, 392
Gulf stream, 158
Gutenberg-Richter law, 118

Habitable zone, 284, 294, 296
Hadean, 17
Haicheng earthquake, 122, 123
Halley's Comet, 80, 378
Hawaii, 104–6
Hayabusa, 70, 71, 81
Heat waves, 144, 145, 196
Heavy water, 229
Heinrich events, 168
Heliosphere, 53, 54, 57
Helium, 57, 245
^{3}He, 229, 304–7
Himalayas, 27, 103, 122, 125, 143, 163, 207
Hinode, 361
HIV/AIDS, 97, 98
Holland, 140, 142, 207
Holocene, 161, 168–71, 201
Hubble Space Telescope, 13, 18, 56, 59, 73, 300, 303
Hurricanes, see cyclones
Huyanaputina, 177
Hybrid cars, 214, 395

Hydrocarbons, conventional, 235
Hydrocarbons, ultimate availability, 235
Hydroelectric power, 213, 217–19, 235
Hydrological cycle, 254
Hydrothermal, 242
Hyperspectral imagery, 350

Ice ages, 42, 161
Ice core, 163
Iceland, 106
ICESAT, 336
India, 5, 124, 141, 143, 228, 266, 322, 341, 378, 382, 38–7, 388, 392, 393, 394, 398
Indium, 222, 244, 249
Indonesia, 103, 104, 106, 108, 126, 266, 406
Indus civilization, 196
Inter Agency Consultative Group (IACG), 378
Interannual variability, 195–7
Interglacials, 153
International Charter on Space and Major Disasters, 379
International cooperation, 306, 369, 370, 378, 379
International Energy Agency, 216, 383
International Space Station, 298
Interstellar travel, 297, 298
Ionosphere, 61, 124, 316, 317, 323, 339, 359–61, 380
IPCC, 2, 188, 309, 310, 321, 322, 372–6, 386–9, 395, 396, 398, 406, 407, 410
  Scenario A1B, 190, 193–6, 208
  Scenario A2, 190, 192, 193, 203, 208, 262, 263
  Scenario B1, 190, 192, 193, 203, 208, 262, 263
  Scenario B2, 190, 192, 193, 203, 208
Iran, 116, 119, 125, 343, 345
Iridium, 43, 44
Iron, 19, 244, 246–8
Irrigation, 267, 270
ITER, 229–32, 250, 304
Itokawa, 70, 71, 81

Japan, 122, 130, 131, 269, 378, 388, 394, 404
Jason 1, 128, 321, 322, 334
Johannesburg World Summit, 391

Kaguya, mission, 300, 302
Katrina, 133, 134, 137, 147, 380, 394
Kilimanjaro, 154–5
Kobe, 121
Krakatoa, 106,107, 109, 110, 126
K-T extinction, 18, 43
Kuiper Belt, 32, 66, 67, 70
Kyoto treaty, 8, 182, 204, 375, 381–96, 408, 410

La Niña, 162
Lagrange point (L_1), 76, 310, 330, 359, 361
Lake Nyos, 46, 108
Lakes, 265
Laki eruption, 46, 110
Landsat, 377
Larsen, captain, 153
Last Glacial Maximum, 168, 169
Last interglacial: Eemian
Late Heavy Bombardment, 17, 18, 21, 22, 35, 62, 289
Less Developed Countries, 266
Limits to Growth, 3–5, 233, 243, 249, 406, 407, 412
Lisbon earthquake, 116, 120, 126
Lithium, 229–31, 249
Lithophile, 241
Little Ice Age, 154, 173, 174, 179, 201
Living with a Star Program (LWS), 362
Low Earth Orbit, 306, 330
Luna missions, 19, 20, 300

Madagascar, 273
Magellan, mission, 19, 283
Magnetars, 61, 62
Magnetic bottle, 229
Magnetosphere, 28, 53–5, 58, 125, 288, 316, 317, 324, 359, 361
MAGSAT, 28, 317, 361
Main Asteroid Belt (MAB), 66–8
Maize, 264, 269
Malaria, 97, 99, 264, 347
Malawi, 271
Maldives, 322
Malthus, Thomas, 34, 264
Marinoan glaciation, 42
Marquis of Pombal, 120
Mars, 13, 14, 19, 21, 31, 34, 66–8, 70, 81, 103, 281–3, 288–93, 297, 298, 300, 303, 308, 310, 311, 405

Mars Express mission, 289, 290
Marshall Plan, 398, 399
Maunder minimum, 174, 176
Maya empires, 196
Medieval Warm Period, 154, 173
Mediterranean region, 103, 194–6, 273
Mercury Planet, 19, 21, 31, 67, 76
Mercury, 249
Mesozoic, 18
Messinian salinity crisis, 138, 143
Meteorites, 14, 16, 31, 62, 64, 70, 76, 241, 294
Methane, 31, 33, 105, 110, 177, 291, 294, 308, 321, 353, 354
Methane hydrates, 233–6
Mg/Ca, 161, 164
Microbial mats, 36
Milankovitch, M., 164, 201
Mineral resources, 6, 242–9, 307
Mitochondrial DNA, 101
Modern humans, 2
MODIS, 355
Molybdenum, 38
Monsoon, 140, 255, 262
Montreal protocol, 5, 8, 182, 325, 384, 385, 393, 396, 408
Moon, 13, 18–21, 31, 62, 63, 229, 281, 282, 289, 291, 293, 294, 295, 297, 300–7, 311,330, 352, 377, 405
Mount Fuji, 109, 345
Mozambique, 140, 142
Mt. St. Helens, 105–7, 110, 355

Natural gas, 232–6
Nature, preservation of, 272, 404
NCAR, 368, 370
NEAR, 64
Near-Earth Asteroids (NEA), 67, 69, 302–7
Near Earth Object (NEO), 68, 74–85, 94
Nenana, 154
Neptune, 13, 67, 296
Nile, 139, 207, 258
Nitrogen fertilizers, 242, 248, 267, 268
Nitrogen oxides, 58, 352, 353, 382
Nitrous oxide, 117
Noachian, 289
Normalized Difference Vegetation Index (NDVI), 329
Northwest Passage, 199
Nuclear energy, 216, 217, 225–8, 235, 237, 301
Nuclear reactors, 225–8

O' Neill, G.K., 299
Obliquity, 166, 167, 291, 292
Ocean tides and waves, 217, 218
Oersted, mission, 28, 317, 361
Oil, 232–6
Oort cloud, 32, 57, 67, 70
Orbiting Carbon Observatory, 324, 353
Ordovician, 161
Ordovician extinction, 43, 61
Ortelius, Abraham, 24
Oxygen, 31, 35–9, 42
  $\delta^{18}O$, 161–4, 169
Ozone hole, 8, 60, 179–82, 351, 376, 407
Ozone layer, 31, 45, 54, 57–8, 60–1, 65, 89, 111–2, 292, 306, 315, 317, 324–6, 351, 358, 373, 384, 385, 386

Pacific Tsunami Warning Center (PTWC), 127
Pakistan, 119, 122, 125, 140, 143, 322, 262
Paleocene-Eocene Thermal Maximum, 203, 236
Paleomagnetism, 26, 27, 30
Paleozoic, 18
Palermo Technical Impact Hazard Scale, 77–9
Palmoil, 224
Pangaea, 24, 25, 27
Path of risk, 84, 85
Permafrost, 168, 207, 289
Permian-Triassic (P-T) extinction, 18, 43, 44, 46
Pesticides, 268
Phanerozoic, 17, 18, 43
Phosphates, 242, 248, 267, 270
Photosynthesis, 223
Phytoplankton, 308, 327, 328, 347
Pinatubo, 104–12, 177, 309, 325
Planetesimals, 16,18, 31, 32
Plasma confinement, 229
Plate tectonics, 13, 19, 23–7, 34, 63, 102, 103, 116, 117, 288, 290, 316, 334
Platinum group, 244, 245, 248
Pleistocene, 18, 60
Pluto, 53,66

Plutonium, 227, 228
Popigai crater, 65
Population, 5, 207, 256, 266, 267
Population growth, 4, 265, 407
Potassium, 242, 248, 268
Precambrian, 17, 25, 27
Precession, 166, 167
Precipitation, 195, 196, 262
Predation, 42, 43
Proterozoic, 17, 23, 35, 37, 38
Pueblo Indians, 196
Pyrite, 38

Quaternary, 18

Radar, 318, 321, 380, 390
  altimetry, 334, 336
  imaging, 64, 66, 283
  interferometry, 124, 319, 342, 343, 345, 346
  synthetic aperture, see SAR
Radiative forcings, 158
Radioactive dating, 14, 15
Radioactivity, 226
Rare Earths, 248
Recycling, 249, 407
Regolith, 69, 70, 292, 301, 303, 304
Relativity theory, 337
Reserve base, 244
Reserves, 244
Reservoir capacity, 256, 260, 263
Resources, 244
Rice, 264, 269
Richter Scale, 117–19
Rio Summit, 375, 384
Rita, 134, 380, 394
River runoff, 254–6, 262
Rodinia, 25, 27
Rosetta probe, 80, 81

Saffir-Simpson's scale, 134
Sagan, Carl, 286, 287
Sahara, 194–7, 205, 268, 272
Sahel, 143, 144, 207, 262, 340, 347
San Francisco earthquake, 119, 122, 125
Saturn, 13, 21, 22, 66, 67, 294, 307
Savannah, 276
Schmitt, Harrison, 304, 305
Sea level, 7, 47, 110, 112, 128, 138, 143, 187, 188, 197–201, 206, 271, 316, 320–22, 336, 394
Sea Surface Temperature, 135, 136, 137, 153, 162, 356
Seismology, 65, 113–16, 122, 130, 317
Shoemaker-Levy 9, 72, 73
Siberia, 233, 258
Siberian traps, 46
Siderophile, 241
Silver, 244, 248
Snow albedo effect, 194
Snowball Earth, 33, 38, 42
SO_2, 382
SOHO satellite, 76, 357, 359, 361
Soil, 268
Solar corona, 174, 176
  cycle, 174, 175, 357
  energy, 214, 221
  energy flux, 191
  forcing, 176, 179, 189, 190
  irradiance, 175, 356, 358, 361
  photovoltaic cells, 217, 221
  proton events, 297
  System, 18, 19, 241
  thermal energy, 217, 222
  ultraviolet, 32, 53, 324, 325, 352, 355, 357–9, 361
  variability, 174, 175, 189
  wind, 53, 54, 174, 176, 318, 359
  X-ray radiation, 360, 361
South Africa, 63, 142, 196, 262, 392, 393
South Atlantic Anomaly, 317
Southwestern USA, 195
Space, cities, 299, 311
  debris, 82, 85–9
  tourism, 293, 301, 306, 311, 412
  Treaty, 306
  weather, 359–61
SPOT satellites, 330, 346, 349, 350
Stardust mission, 72
Stern review, 207, 389, 407
Stomatolites, 37
Storegga slide, 126
Stratosphere, 61, 109, 110, 285, 309, 310, 316, 323, 324, 326, 352, 357, 373, 376, 384
Strontium, 15
Sugar cane, 224
Sumatra, 106–8, 126

Sumatra-Andaman tsunami, 94, 118–24, 127–9, 131, 134, 147, 394
Summer ice, 199
Sun-synchronous orbit, 330
Supernova, 59–61, 240
Sustainable development, 3, 384
Swarm mission, 318, 319, 361
Switchgrass, 224
Synthetic Aperture Radars (SARs), 339–49

Tambora, 106–10, 176, 177
Tangshan earthquake, 119, 122, 123
Tank to wheel efficiencies, 214
Temperature, global, 153, 172, 177, 193–5
  land, 170, 194
  past, 34, 153, 154, 161–73
  variability, 195–7
Temple 1, 72
Terra preta, 272
Terraforming, 282, 285–8, 291–4, 296, 308
TerraSAR, 340
Tertiary, 18
Thermohaline circulation, 158, 159, 168, 169
Thetys Sea, 234
Thorium, 228, 245, 249
Three Gorges Dam, 142, 219, 258
Titan, 19, 103, 281–3, 294, 308, 311
Toba, 106–10
Tokamak, 230
Topex-Poseidon, 128, 321, 322, 334
Torino Scale, 78–80
Tree of Life, 36, 37
Tree rings 171, 173
Triassic mass extinction, 18, 43, 47
Tritium, 226, 229–32
Tropical Rainfall Measurement Mission (TRMM), 380
Troposphere, 109, 110, 285, 309, 310, 323, 324, 326, 339, 352, 357, 358
Tsiolkovsky, K., 281, 298, 299, 311
Tsunamis, 44, 74, 93, 94, 102, 108, 109, 116, 119, 120, 125–31, 134, 138, 147, 309, 338, 369, 393, 394
Tundra, 265, 272, 324
Tunguska, 73
Turkey, 116, 122, 124, 379

Ultraviolet (see also solar ultraviolet), 60, 61, 285, 351, 352, 355,
UN Environment Program (UNEP), 374–6, 298, 409
UN Framework Convention on Climate Change (UNFCCC), 375, 376, 384, 385, 388, 395
UNESCO, 131, 375
United Nations, 10, 62, 85, 87, 94, 311, 371, 374, 376, 382, 387, 391–3, 397, 398, 403, 408–10
United Nations Development Program (UNDP), 96, 148
Uranium, 225–8
Uranium, oceanic, 227
Uranus, 13, 286, 306, 307, 308

Van Allen radiation belts, 28, 359
Venezuela, 233
Venice, 139, 140, 141, 346
Venus, 13, 19, 21, 31, 33, 103, 281–7, 308, 319
Venus Express mission, 285
Vesuvius, 114
Vinland, 154
Virtual water trade, 260
Volcanic Explosivity Index (VEI), 106–9, 112
Volcanism, 19, 32, 46, 102–15, 176, 177, 316, 325, 355
Vostok ice core, 163

Water, 6, 223, 253–63, 270, 284, 286–9, 319–23, 333, 403, 409, 410
  atmospheric, 339
  cycle, 316, 327
  ice, 290, 291, 294, 295
  management, 408
  origin, 32
  pollution, 253
  quality, 382
  resources, 255, 316, 376, 390, 407
  runoff, 255, 256
  stress, 257
  vapor, 285, 354, 356, 380
  waves, 125
  withdrawals, 257, 256, 260
Weather forecasting, 316, 368, 369, 380

Weathering, 33
Wegener, Alfred, 24, 27
West Antarctic Ice Sheet, 155, 170, 187, 199–201, 204
Wheat, 264, 270
Wind energy, 217, 219–21, 235, 237
WMAP, 13
World Health Organization (WHO), 95, 97, 98, 99, 139, 143, 147, 352
World Meteorological Organization (WMO), 132, 141, 369, 370–2, 374, 375, 378, 379, 392, 397, 410
WRE scenarios, 192, 193

Yangtse, 143, 258, 263
Yellow river, 143, 258
Younger Dryas, 168, 169, 205, 263, 272
Yucca mountain, 2, 227

Zinc, 244, 245
Zircon, 15, 16, 33