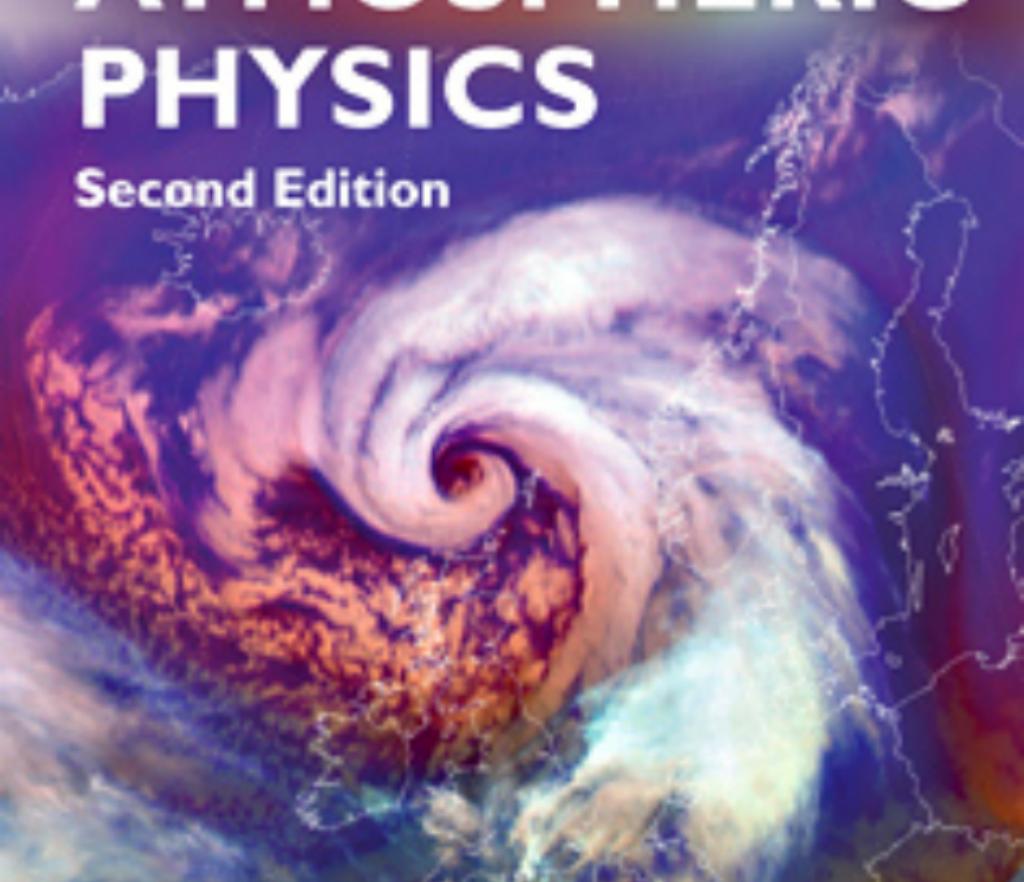


DAVID G. ANDREWS

An Introduction to ATMOSPHERIC PHYSICS

Second Edition



CAMBRIDGE

www.cambridge.org/9780521872201

This page intentionally left blank

An Introduction to Atmospheric Physics

Second Edition

The new edition of David Andrews' excellent textbook has been significantly revised and expanded. This textbook provides a quantitative introduction to the Earth's atmosphere for intermediate-advanced undergraduate and graduate students. It now includes a new chapter on the physics of climate change which builds upon material introduced in earlier chapters, giving the student a broad understanding of some of the physical concepts underlying this most important and topical subject. In contrast to many other books on atmospheric science, the emphasis is on the underlying physics. Atmospheric applications are developed mainly in the problems given at the end of each chapter. The book is an essential resource for all students of atmospheric physics as part of an atmospheric science, meteorology, physics, Earth science, planetary science, or applied mathematics course.

David Andrews has been a lecturer in physics at Oxford University and a physics tutor at Lady Margaret Hall, Oxford, for 20 years. During this time he has had extensive experience of teaching a wide range of physics courses, including atmospheric physics. This experience has included giving lectures to large student audiences and also giving tutorials to small groups. Tutorials, in particular, have given him insights into the problems that physics students encounter when learning atmospheric physics, and the kinds of topic that excite them. His broad teaching experience has also helped him to introduce students to connections between topics in atmospheric physics and related topics in other areas of physics. He feels that it is particularly important to expose today's physics students to the stimulation and challenges presented by the atmosphere and climate. He has also published a graduate textbook, *Middle Atmosphere Dynamics*, with J. R. Holton and C. B. Leovy (1987, Academic Press). He is a fellow of the Royal Meteorological Society and a member of the Institute of Physics.

An Introduction to Atmospheric Physics

Second Edition

DAVID G. ANDREWS



CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521872201

© D. G. Andrews 2010
First edition © Cambridge University Press 2000

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2010

ISBN-13 978-0-511-72966-9 eBook (NetLibrary)

ISBN-13 978-0-521-87220-1 Hardback

ISBN-13 978-0-521-69318-9 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>Preface to the Second Edition</i>	<i>page</i> ix
1 Introduction	
1.1 The atmosphere as a physical system	1
1.2 Atmospheric models	3
1.3 Two simple atmospheric models	4
1.4 Some atmospheric observations	7
1.5 Weather and climate	17
Further reading	18
2 Atmospheric thermodynamics	
2.1 The ideal gas law	19
2.2 Atmospheric composition	20
2.3 Hydrostatic balance	22
2.4 Entropy and potential temperature	24
2.5 Parcel concepts	26
2.6 The available potential energy	30
2.7 Moisture in the atmosphere	32
2.8 The saturated adiabatic lapse rate	37
2.9 The tephigram	40
2.10 Cloud formation	42
Further reading	48
Problems	48
3 Atmospheric radiation	
3.1 Basic physical concepts	52
3.2 The radiative-transfer equation	57
3.3 Basic spectroscopy of molecules	63
3.4 Transmittance	69
3.5 Absorption by atmospheric gases	71
3.6 Heating rates	75
3.7 The greenhouse effect revisited	81
3.8 A simple model of scattering	86
Further reading	88
Problems	89

4 Basic fluid dynamics	94
4.1 Mass conservation	94
4.2 The material derivative	95
4.3 An alternative form of the continuity equation	98
4.4 The equation of state for the atmosphere	99
4.5 The Navier–Stokes equation	99
4.6 Rotating frames of reference	102
4.7 Equations of motion in coordinate form	104
4.8 Geostrophic and hydrostatic approximations	107
4.9 Pressure coordinates and geopotential	111
4.10 The thermodynamic energy equation	113
Further reading	114
Problems	114
5 Further atmospheric fluid dynamics	119
5.1 Vorticity and potential vorticity	119
5.2 The Boussinesq approximation	122
5.3 Quasi-geostrophic motion	125
5.4 Gravity waves	128
5.5 Rossby waves	132
5.6 Boundary layers	136
5.7 Instability	141
Further reading	147
Problems	147
6 Stratospheric chemistry	151
6.1 Thermodynamics of chemical reactions	151
6.2 Chemical kinetics	153
6.3 Bimolecular reactions	155
6.4 Photo-dissociation	157
6.5 Stratospheric ozone	158
6.6 The transport of chemicals	161
6.7 The Antarctic ozone hole	164
Further reading	168
Problems	168
7 Atmospheric remote sounding	171
7.1 Atmospheric observations	171
7.2 Atmospheric remote sounding from space	172
7.3 Atmospheric remote sounding from the ground	182
Further reading	189
Problems	190

8 Climate change	195
8.1 Introduction	195
8.2 An energy balance model	198
8.3 Some solutions of the linearised energy balance model	200
8.4 Climate feedbacks	204
8.5 The radiative forcing due to an increase in carbon dioxide	207
Further reading	211
Problems	212
9 Atmospheric modelling	215
9.1 The hierarchy of models	215
9.2 Numerical methods	217
9.3 Uses of complex numerical models	219
9.4 Laboratory models	220
9.5 Final remarks	222
Further reading	224
<i>Appendix A Useful physical constants</i>	225
<i>Appendix B Derivation of the equations of motion in spherical coordinates</i>	227
<i>References</i>	229
<i>Index</i>	234

Preface to the Second Edition

Atmospheric physics has a long history as a serious scientific discipline, extending back at least as far as the late seventeenth century. Today it is a rich and fascinating subject, sustained by detailed global observations and underpinned by solid theoretical foundations. It provides an essential tool for tackling a wide range of environmental questions, on local, regional and global scales. Although the solutions to vital and challenging problems concerning weather forecasting and climate prediction rely heavily on the use of supercomputers, they rely even more on the imaginative application of soundly based physical insights.

This book is intended as an introductory working text for third- or fourth-year undergraduates studying atmospheric physics as part of a physics, meteorology, or earth and planetary sciences degree course. It should also be useful for graduate students who are studying atmospheric physics for the first time and for students of applied mathematics, physical chemistry and engineering who have an interest in the atmosphere. Physics undergraduates, in particular, will discover that a sound understanding of atmospheric physics can be built up in the same quantitative and logical manner as the other areas of physics that they encounter in their courses.

Modern scientific study of the atmosphere draws on many branches of physics. I believe that a balanced introductory course in atmospheric physics should include at least some atmospheric thermodynamics, radiative transfer, atmospheric fluid dynamics and elementary atmospheric chemistry. Armed with a basic understanding of these topics, the interested student will be able to grasp the essential physics behind important issues of current concern – such as how the climate changes in response to increases in greenhouse gases, and why depletion of stratospheric ozone has occurred – as well as more familiar processes such as the formation of raindrops and the development of weather systems.

This book therefore aims to show how basic physical principles can be applied to help us to understand the workings of the Earth's atmosphere. It includes treatments of the topics mentioned in the previous paragraph, plus a few others. Attention is restricted to the troposphere, stratosphere and mesosphere, that is, the region between the ground and about 90 km altitude. Although other planets are seldom mentioned explicitly, many of the topics covered also apply to the atmospheres of Venus and Mars.

In contrast to many other books on atmospheric science, the emphasis in the text is on the *underlying physics*; atmospheric applications are developed mainly in the problems given at the end of each chapter. It is essential that the serious student should attempt some of these problems, to test his or her understanding of the material and to obtain a broader perspective on the subject than can be provided by the text alone. (In some cases important meteorological applications have been omitted because they rely on semi-empirical rules

rather than on basic physics; there are excellent meteorology books covering this type of material.) Solutions to the problems are provided on a password-protected website for the benefit of course instructors.

The book assumes a basic knowledge of thermodynamics, electromagnetic radiation and quantum physics, together with some elementary vector calculus, at about the level reached in core physics courses at universities. It does not assume prior knowledge of fluid dynamics – which is frequently (and I believe mistakenly) omitted from core physics courses. Most of the material included here is based on over twenty years' experience teaching atmospheric physics to undergraduate physicists at Oxford University.

This Second Edition includes a new chapter on the physics of climate change, which builds upon material introduced in earlier chapters and aims at giving the student a broad understanding of some of the physical concepts underlying this most important and topical subject. I have also corrected and updated several other chapters, figures and problems.

Course instructors can use the book in its entirety, or can select topics of particular interest to them. However, I would strongly recommend covering most sections of Chapters 2 (thermodynamics), 3 (radiation) and 4 (basic fluid dynamics) as a minimum. Later chapters depend on these three in various ways: for example, Chapter 5 depends heavily on Chapter 4, Chapter 6 requires a little knowledge of Chapter 2, and Chapters 7 and 8 require a good understanding of parts of Chapter 3.

Several colleagues have provided invaluable assistance with this new edition, especially with the new Chapter 8. In particular, Myles Allen has shared many stimulating ideas with me on how the physics of climate change should be taught, though I take full responsibility for any shortcomings in my treatment of the subject. I would also like to acknowledge help and advice from Stephen Blundell, Anu Duhia, Jonathan Gregory, William Ingram, Guy Peskett, Keith Shine and Philip Stier, and I thank all those colleagues and students who drew my attention to errors in the First Edition. Once again my wife, Kathleen Daly, gave much support and encouragement during the writing process.

The following publishers have kindly given permission to reproduce or adapt figures:

The American Meteorological Society: [Figure 1.7](#).

Springer Science and Business Media: [Figures 3.15 and 3.17](#).

Pearson Education, Inc., Upper Saddle River, New Jersey: [Figure 3.17](#).

Oxford University Press: [Figures 5.1, 5.2 and 6.7](#).

The Royal Meteorological Society: [Figure 6.6](#).

The Optical Society of America: [Figure 7.8](#).

Elsevier: [Figure 7.16](#).

Intergovernmental Panel on Climate Change: [Figure 8.1](#)

Taylor and Francis Ltd (<http://www.informaworld.com>): [Figure 9.4](#).

I also thank the authors of these figures for permission to reproduce or adapt their material.

[Figure 6.5](#) is drawn using data provided by J. D. Shanklin, British Antarctic Survey, Madingley Road, Cambridge, England, CB3 0ET.

This chapter gives a quick sketch of some of the material to be covered in this book. We start in [Section 1.1](#) with an outline of some of the more important physical processes that occur in the Earth's atmosphere. To interpret atmospheric observations we need to develop physical and mathematical models; they are briefly discussed in [Section 1.2](#). Two extremely simple models are introduced in [Section 1.3](#): the second of these includes a very basic representation of the greenhouse effect. In [Section 1.4](#) we present a selection of observations of atmospheric processes, together with simple physical explanations for some of them. In [Section 1.5](#) we briefly mention some ideas on weather and climate.

1.1 The atmosphere as a physical system

The Earth's atmosphere is a natural laboratory, in which a wide variety of physical processes takes place. The purpose of this book is to show how basic physical principles can help us model, interpret and predict some of these processes. This section presents a brief overview of the physics involved.

The atmosphere consists of a mixture of ideal gases: although molecular nitrogen and molecular oxygen predominate by volume, the minor constituents carbon dioxide, water vapour and ozone play crucial roles. The forcing of the atmosphere is primarily from the Sun, though interactions with the land and the ocean are also important.

The atmosphere is continually bombarded by solar photons at infra-red, visible and ultra-violet wavelengths. Some solar photons are scattered back to space by atmospheric gases or reflected back to space by clouds or the Earth's surface; some are absorbed by atmospheric molecules (especially water vapour and ozone) or clouds, leading to heating of parts of the atmosphere; and some reach the Earth's surface and heat it. Atmospheric gases (especially carbon dioxide, water vapour and ozone), clouds and the Earth's surface also emit and absorb infra-red photons, leading to further heat transfer between one region and another, or loss of heat to space. Some of these **radiative-transfer** processes are discussed in [Chapter 3](#). Solar photons may also be energetic enough to disrupt molecular chemical bonds, leading to photochemical reactions; see [Chapter 6](#).

The atmosphere is generally close to hydrostatic balance in the vertical, except on small scales; that is, the weight of each horizontal slab of atmosphere is supported by the difference in pressure between its lower and upper surfaces. An alternative statement of

this physical fact is that there is a balance between vertical pressure gradients and the gravitational force per unit volume acting on each portion of the atmosphere. On combining the equation describing hydrostatic balance with the ideal gas law we find that, in a hypothetical isothermal atmosphere, the pressure and density would fall exponentially with altitude (see [Section 2.3](#)). In the real, non-isothermal, atmosphere the pressure and density variations are usually still close to this exponential form, with an *e*-folding height of about 7 or 8 km. Gravity thus tends to produce a **density stratification** in the atmosphere.

Given a density stratification of this kind, a small portion of air that is displaced upwards from its equilibrium position will be negatively buoyant compared with its surroundings and hence will fall back towards equilibrium, under gravity; similarly a downward-displaced portion will rise back towards its equilibrium position. Buoyancy therefore acts as a restoring force; the atmosphere is said to be **stably stratified**. The strength of the stability of the stratification varies from one part of the atmosphere to another.

Thermodynamic principles are essential for describing many atmospheric processes. For example, any consideration of the effects of atmospheric heating or cooling will make use of the First Law of Thermodynamics. The concept of entropy (or the closely related quantity, potential temperature) frequently assists interpretation of atmospheric behaviour. Knowledge of changes in phase between vapour, liquid and solid forms of the water in the atmosphere is crucial for an understanding of the formation of rain and snow. Moreover, the associated latent heating and cooling can provide important contributions to heat transfer within the atmosphere–ocean system; for example, evaporation of a droplet of sea water at one location and subsequent condensation of the resulting water vapour into a droplet at another location in the atmosphere transfers heat from the ocean to the atmosphere. The basics of atmospheric thermodynamics are covered in [Chapter 2](#).

In atmospheric physics we use the usual macroscopic definitions of the temperature and pressure of a gas. From the kinetic theory of gases, these have well-known interpretations in terms of the mean kinetic energy of molecules and the mean transfer of momentum by molecules, respectively. When considering dynamical processes – that is, the response of atmospheric motions to applied forces – we can average other physical quantities such as density and velocity over many molecules and regard the atmosphere as a continuous fluid; individual molecular motions need not be taken into account. It is clear from the most cursory weather observations that the resulting bulk fluid motion of the atmosphere is still very complex. However, when the motion is viewed on a large scale (say hundreds of kilometres in horizontal extent), some simplifying features appear. In particular, Coriolis forces play significant roles: these forces result from the rotation of the Earth and tend to deflect a moving portion of air to the right of its motion in the Northern Hemisphere and to the left in the Southern Hemisphere. A near balance between Coriolis forces and horizontal pressure gradient forces leads to wind motions that circulate along isobars (surfaces of constant pressure) at a given height. The sense of the circulation is anticlockwise around low-pressure regions and clockwise around high-pressure regions in the Northern Hemisphere and vice versa in the Southern Hemisphere. The basic principles of atmospheric fluid dynamics are introduced in [Chapter 4](#).

An important feature of the buoyancy restoring effect in a stably stratified atmosphere is that it can support fluid-dynamical waves, known as **gravity waves**,¹ in which the fluid pressure, density, temperature and velocity fluctuate together. These waves may propagate, allowing one part of the atmosphere to ‘communicate’ over great distances with other parts, without a corresponding transport of mass. The Coriolis force can also act as a restoring force, giving rise to further types of fluid wave motion. In particular, on large scales we find **Rossby waves**, which depend crucially on the rotation and the spherical geometry of the Earth and are associated with many observed large-scale disturbances in the troposphere and the stratosphere. As in many other branches of physics, the study of wave motions is an essential part of atmospheric physics; see [Chapter 5](#).

As noted above, solar radiation can initiate photochemical reactions by dissociating atmospheric molecules. A host of other types of chemical reaction between atmospheric molecules, both natural and man-made, can also occur. Atmospheric chemistry is a large and highly complex subject; in this book we focus on one small but significant branch of the subject, namely the chemistry of stratospheric ozone. This provides a good example of physical principles in operation and is highly topical, with direct application to the Antarctic ozone hole and global depletion of ozone; see [Chapter 6](#). It also demonstrates the importance of atmospheric transport processes, by which the winds blow chemical species from one part of the globe to another.

No study of the atmosphere can make progress without suitable observations, and all observational techniques rely to some extent on physical principles. One important type of observational technique is that of **remote sounding**, which depends on the detection of electromagnetic radiation emitted, scattered or transmitted by the atmosphere. In [Chapter 7](#) we describe several examples of remote sounding, looking both at space-borne and at ground-based methods.

Climate change is a topic of great current concern. An understanding of how the Earth’s climate has changed in the past, what determines its current state, and how it will change in the future depends on a detailed knowledge of a wide variety of physical processes, some of which are touched upon in this book. An outline of some of the more important physical concepts and processes associated with climate change is given in [Chapter 8](#).

1.2 Atmospheric models

Unlike laboratory physicists, atmospheric researchers cannot perform controlled experiments on the large-scale atmosphere. The standard ‘scientific method’, of observing phenomena, formulating hypotheses, testing them by experiment, then formulating revised hypotheses and so on, cannot be applied directly. Instead, after an atmospheric phenomenon is discovered, perhaps by sifting through a great deal of data, we develop

¹ Not to be confused with the gravitational waves of General Relativity!

models, which incorporate representations of those processes that we hypothesise are most important for causing the phenomenon. Models act as surrogate atmospheres, on which ‘thought experiments’ can be performed. These models are usually formulated in terms of mathematical equations, and the ‘experiments’ are performed by solving these equations (perhaps by computer) under various conditions and interpreting the solutions in terms of physical behaviour. Occasionally a laboratory apparatus may provide a useful atmospheric model. The performance of the model (and thus the appropriateness of the hypothesised set of processes) is judged by comparing the model’s behaviour with that of the atmosphere.

Normally, a **hierarchy** of models is used, starting with simple ‘back-of-the-envelope’ models and progressing through models of intermediate complexity to the highly complex ‘general circulation models’ which require large computer resources. The models considered in this book are mostly of the simpler type, although the more complex ones are briefly discussed in [Chapter 9](#). Since the simpler models can usually be investigated analytically and their workings fully explored, they can provide a basic ‘physical intuition’, which can then be applied to the interpretation of the more complex models. Because of their very simplicity, however, they cannot usually be expected to give accurate simulations of observed atmospheric behaviour.

The more comprehensive models bring together many of the physical principles introduced in this book and allow for interactions between them. Some of these interactions may involve complex feedbacks, however, so it may be difficult to establish causal relationships among the various processes that are involved.

Historically, a major use of comprehensive atmospheric models has been for weather forecasting. With the development of supercomputers, reliable longer-term climate forecasting is also becoming a feasible proposition, although the models used for this purpose are still inadequate in some respects. Complex models are also used for **data assimilation**, by providing a dynamically self-consistent means of interpolating, in space and time, between sparse data points from a variety of sources. Data assimilation aims to provide an accurate estimate of the time-evolving, three-dimensional state of the atmosphere, and is nowadays a vital component of the weather-forecasting process.

1.3 Two simple atmospheric models

It is a basic observational fact that the Earth’s mean surface temperature is about 288 K. In this section we consider whether this can be explained in simple terms, given the input of solar radiation and some elementary atmospheric physics. We consider two models; the first turns out to be seriously defective, but the second, which includes a simple representation of the **greenhouse effect**, gives a surface temperature in reasonable agreement with observations. Both models assume that radiation is the only heat-transfer process. To quantify this process we introduce the **irradiance**, i.e. the power per unit area, associated with any given stream of radiation: see [Section 3.2.1](#) for a more precise definition.

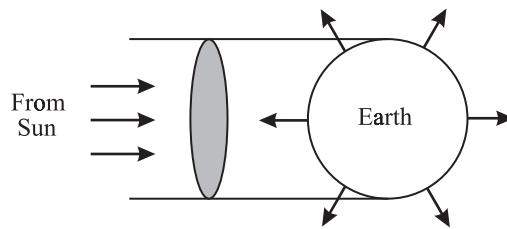


Fig. 1.1

Illustrating the calculation of the temperature of the Earth, ignoring any absorption of radiation by the atmosphere. The parallel arrows indicate solar radiation, confined within a tube of cross-sectional area πa^2 . The radial arrows indicate outgoing thermal radiation from the total surface area $4\pi a^2$ of the Earth.

1.3.1 A model with a non-absorbing atmosphere

The solar power per unit area at the Earth's mean distance from the Sun (the **total solar irradiance**, TSI, formerly called the **solar constant**) is $F_s = 1370 \text{ W m}^{-2}$. The solar beam is essentially parallel at the Earth, so the power that is intercepted by the Earth is contained in a tube of cross-sectional area πa^2 , where a is the Earth's radius; see Figure 1.1. The total solar energy received per unit time is therefore $F_s \pi a^2$.

We assume that the Earth–atmosphere system has a **planetary albedo** A equal to 0.3; that is, 30% of the incoming solar radiation is reflected back to space without being absorbed: this is close to the observed mean value. The Earth therefore reflects $0.3F_s \pi a^2$ of the incoming solar power back to space.

If the Earth is assumed to emit as a **black body** at a uniform absolute temperature T then, by the Stefan–Boltzmann law,

$$\text{Power emitted per unit area} = \sigma T^4, \quad (1.1)$$

where σ is the Stefan–Boltzmann constant.² However, power is emitted in all directions from a total surface area $4\pi a^2$, so the total power emitted is $4\pi a^2 \sigma T^4$. We assume in the present model that all of this power is transmitted to space, with none absorbed by the atmosphere. Then, assuming that the Earth is in thermal equilibrium, the incoming and outgoing power must balance, so

$$(1 - A)F_s \pi a^2 = 4\pi a^2 \sigma T^4. \quad (1.2)$$

On substituting the values of A and F_s into this, we obtain $T \approx 255 \text{ K}$. The temperature obtained from this calculation is called the **effective emitting temperature** of the Earth: see equation (3.36). Its value is significantly lower than the observed mean surface temperature of about 288 K. The present model is clearly lacking in some vital ingredient; we find in Section 1.3.2 that inclusion of the radiation-trapping effect of the atmosphere (the ‘greenhouse effect’) leads to a surface temperature that is much closer to reality.

² The concept of a black body is explained in Section 3.1.1; for the moment, all that is required is that the power per unit area emitted by a black body satisfies equation (1.1). The value of σ is given in Appendix A, together with the values of other useful physical constants.

1.3.2 A simple model of the greenhouse effect

We now consider the effect of adding a layer of atmosphere, of uniform temperature T_a , to the model of Section 1.3.1; see Figure 1.2. The atmosphere is assumed to transmit a fraction \mathcal{T}_{sw} of any incident solar (short-wave) radiation and a fraction \mathcal{T}_{lw} of any incident thermal (infra-red, or long-wave) radiation (these fractions are called **transmittances**: see Section 3.4), and to absorb the remainder. We assume that the ground is at temperature T_g .

Taking account of albedo effects and the difference between the area of the emitting surface $4\pi a^2$ and the intercepted cross-sectional area πa^2 of the solar beam (see Section 1.3.1), the mean unreflected incoming solar irradiance at the top of the atmosphere is

$$F_0 = \frac{1}{4}(1 - A)F_s, \quad (1.3)$$

or about 240 W m^{-2} with the given values of A and F_s . Of this, an amount $\mathcal{T}_{\text{sw}}F_0$ is absorbed by the ground and the remainder $(1 - \mathcal{T}_{\text{sw}})F_0$ is absorbed by the atmosphere.

The ground is assumed to emit as a black body. By equation (1.1) it therefore emits an upward irradiance $F_g = \sigma T_g^4$, of which a proportion $\mathcal{T}_{\text{lw}}F_g$ reaches the top of the atmosphere, the remainder being absorbed by the atmosphere. The atmosphere is not a black body, but emits irradiances $F_a = (1 - \mathcal{T}_{\text{lw}})\sigma T_a^4$ both upwards and downwards, as shown in Figure 1.2. (By Kirchhoff's law, the emittance – the ratio of the actual emitted irradiance to the irradiance that would be emitted by a black body at the same temperature – equals the absorptance $1 - \mathcal{T}_{\text{lw}}$; see Section 3.1.1.)

We now assume that the system is in **radiative equilibrium**: that is, energy transfer takes place only by the radiative processes described above, and the associated irradiances are in balance everywhere; we neglect any energy transfers due to non-radiative processes such as fluid motions. Equating irradiances, we have

$$F_0 = F_a + \mathcal{T}_{\text{lw}}F_g \quad (1.4a)$$

above the atmosphere, and

$$F_g = F_a + \mathcal{T}_{\text{sw}}F_0 \quad (1.4b)$$

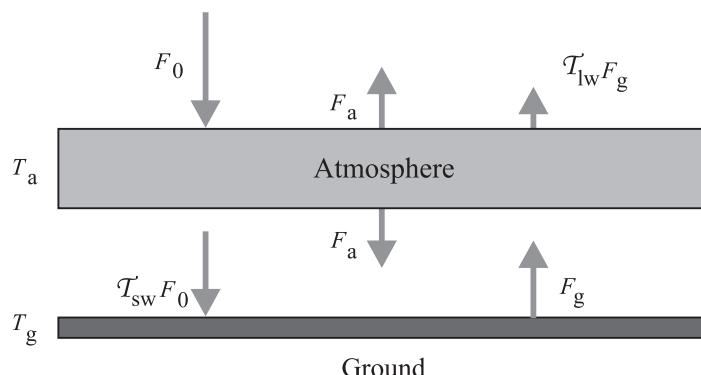


Fig. 1.2

A simple model of the greenhouse effect. The atmosphere is taken to be a layer at temperature T_a and the ground a black body at temperature T_g . Various solar and thermal irradiances are shown.

between the atmosphere and the ground. By eliminating F_a from equations (1.4) we obtain

$$F_g = \sigma T_g^4 = F_0 \frac{1 + \mathcal{T}_{sw}}{1 + \mathcal{T}_{lw}}. \quad (1.5)$$

In the absence of an absorbing atmosphere, we would have $\mathcal{T}_{sw} = \mathcal{T}_{lw} = 1$, so F_g would equal F_0 , giving $T_g \approx 255\text{ K}$, as in Section 1.3.1. Taking rough values for the Earth's atmosphere to be $\mathcal{T}_{sw} = 0.9$ (strong transmittance and weak absorption of solar radiation) and $\mathcal{T}_{lw} = 0.2$ (weak transmittance and strong absorption of thermal radiation), we obtain a surface temperature of $T_g \approx 286\text{ K}$, which is quite close to the observed mean value of 288 K. This close agreement is somewhat fortuitous, however, since in reality non-radiative processes also contribute significantly to the energy balance.

We can also find the atmospheric emission from equations (1.4):

$$F_a = (1 - \mathcal{T}_{lw})\sigma T_a^4 = F_0 \frac{1 - \mathcal{T}_{sw}\mathcal{T}_{lw}}{1 + \mathcal{T}_{lw}} \quad (1.6)$$

and this gives the temperature of the model atmosphere, $T_a \approx 245\text{ K}$.

This model provides a simple example of the **greenhouse effect**: the raised surface temperature is due to the fact that there is less absorption (greater transmittance) for solar radiation than there is for thermal radiation. Thus the atmosphere readily transmits solar radiation but tends to trap thermal radiation.³ Atmospheric gases that absorb and emit infra-red radiation but allow solar radiation to pass through relatively unscathed are called **greenhouse gases**.

One way to quantify the 'greenhouse effect' of an absorbing gas is in terms of the amount $F_g - F_0$ by which it reduces the outgoing irradiance from its surface value: in the case discussed above this reduction is 140 W m^{-2} . This equals the difference between the amount $(1 - \mathcal{T}_{lw})\sigma T_g^4 = 304\text{ W m}^{-2}$ of the thermal emission from the 'warm' surface that is absorbed by the 'cool' atmosphere and the smaller amount $F_a = (1 - \mathcal{T}_{lw})\sigma T_a^4 = 164\text{ W m}^{-2}$ that the atmosphere re-emits upwards. Since the atmosphere is in equilibrium, it also equals the difference between the downward emission F_a from the atmosphere and the small proportion $(1 - \mathcal{T}_{sw})F_0 = 24\text{ W m}^{-2}$ of the solar irradiance that it absorbs.

1.4 Some atmospheric observations

In this section we present a selection of examples of basic atmospheric observations and give some indication of their physical explanation. Further details are given in later chapters of this book.

³ The term 'greenhouse effect' is a misnomer, however, since the elevated temperature in a greenhouse does not primarily depend on the similar radiative properties of glass, but rather on the suppression of convective heat loss.

1.4.1 The mean temperature and wind fields

Figure 1.3 shows a typical example of the vertical structure of the temperature in the lowest 100 km of the atmosphere. The atmosphere is conventionally divided into layers in the vertical direction, according to the variation of temperature with height. The layer from the ground up to about 15 km altitude, in which the temperature decreases with height, is called the **troposphere** and is bounded above by the **tropopause**. The layer from the tropopause to about 50 km altitude, in which the temperature rises with altitude, is called the **stratosphere** and is bounded above by the **stratopause**. The layer from the stratopause to about 85–90 km, in which the temperature again falls with altitude, is called the **mesosphere** and is bounded above by the **mesopause**. Above the mesopause is the **thermosphere**, in which the temperature again rises with altitude.

The troposphere is also called the **lower atmosphere**. It is here that most ‘weather’ phenomena, such as cyclones, fronts, hurricanes, rain, snow, thunder and lightning, occur.

The stratosphere and mesosphere together are called the **middle atmosphere**. A notable feature of the stratosphere is that it contains the bulk of the ozone molecules in the atmosphere; see Figure 1.4. The neighbourhood of the ozone maximum in the lower stratosphere is loosely known as the **ozone layer**. The production of ozone (O_3) molecules occurs through photochemical processes involving the absorption of solar ultra-violet photons by molecular oxygen (O_2) in the stratosphere, three O_2 molecules eventually forming two O_3 molecules. The equilibrium profile of ozone depends also on chemical ozone-destruction processes and on the transport of ozone by the winds (see Chapter 6).

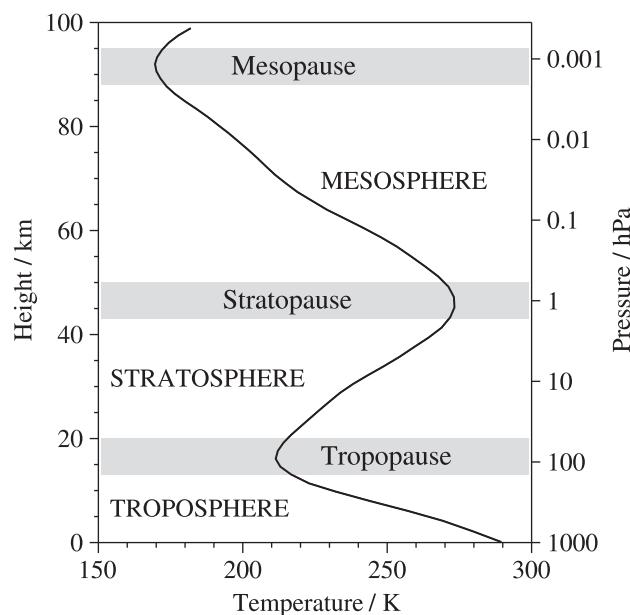


Fig. 1.3

Typical vertical structure of atmospheric temperature (K) in the lowest 100 km of the atmosphere. Based on data from Fleming *et al.* (1990).

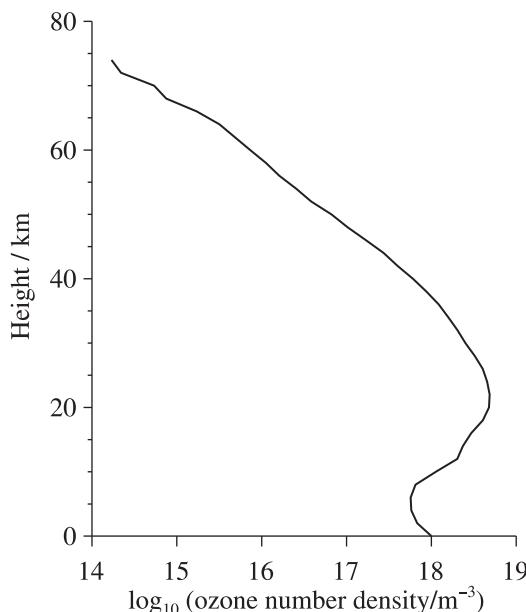


Fig. 1.4 Typical vertical structure of the mean midlatitude ozone number density (molecules m^{-3}). Based on data from US Standard Atmosphere (1976).

Above the middle atmosphere is the **upper atmosphere**, where effects of ionisation become dominant in determining the atmospheric structure and the air becomes so rarefied that the assumption that it can be treated as a continuous fluid starts to break down. In this book we concentrate on the physics of the lower and middle atmospheres.

From hydrostatic balance, the pressure at any level in the atmosphere is proportional to the mass of air above that level. From the pressure axis in Figure 1.3 it follows that approximately 90% of the atmospheric mass is in the troposphere, a little under 10% in the stratosphere and only about 0.1% in the mesosphere and above.⁴ Despite their relatively low mass, the stratosphere and mesosphere are not insignificant, however. For example, ozone in the stratosphere absorbs ultra-violet solar radiation, thereby protecting the biosphere from potentially damaging effects.

Figure 1.3 is not representative of all latitudes and seasons. A more comprehensive plot, of the zonal-mean (i.e. longitudinally averaged) temperature averaged over several Januaries, as a function of height and latitude, is given in Figure 1.5. It will be seen that although the general shape of the vertical variation of temperature in midlatitudes is roughly in accord with that in Figure 1.3, there are significant latitudinal variations of the heights and magnitudes of the temperature extrema. For example, the equatorial tropopause is at a greater altitude and colder than that at higher latitudes, the summer stratopause is lower

⁴ In this book we use the unit hPa for pressure ($1 \text{ hPa} = 10^2 \text{ Pa}$). This is equivalent to the millibar, which was formerly in common use in meteorology.

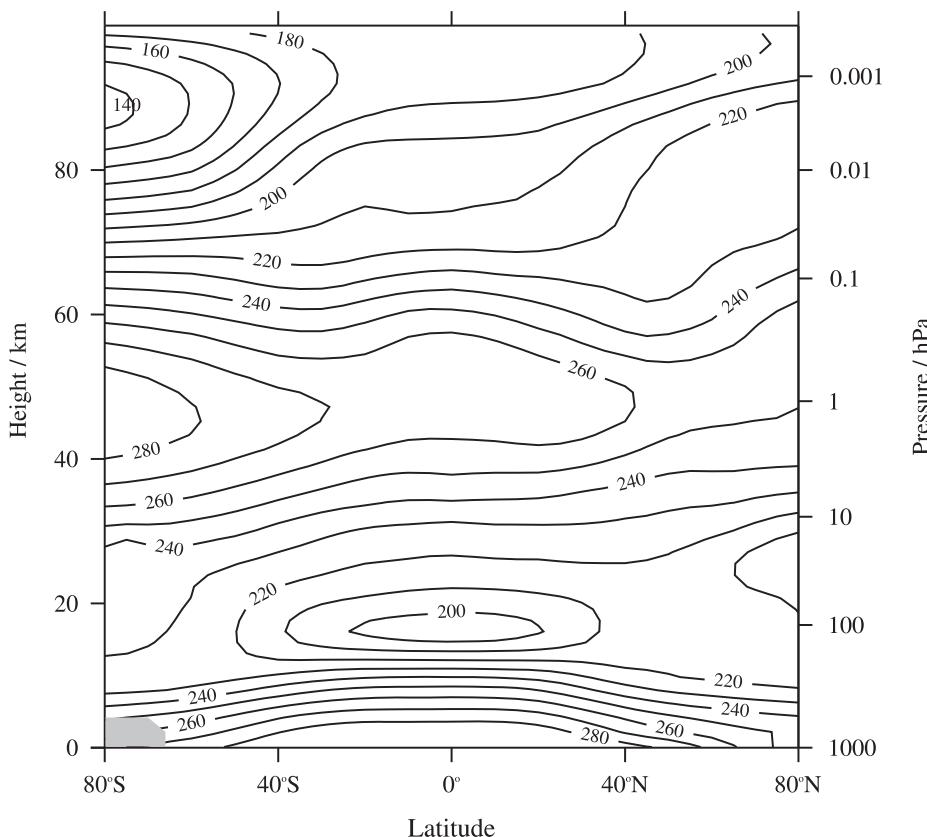


Fig. 1.5 Zonal-mean temperature (K) for January, from the CIRA (COSPAR International Reference Atmosphere) dataset. A small region at low levels over Antarctica is omitted. Based on data from Fleming *et al.* (1990).

and warmer than the winter stratopause and the summer mesopause is extremely cold. (In fact the lowest natural terrestrial temperatures are found there.)

Some of these temperature features can be crudely explained in terms of simple physical mechanisms. For example, the warm stratopause can be attributed to the ozone distribution, which peaks in the stratosphere; absorption of solar radiation by the ozone leads to heating of the upper stratosphere and, since in equilibrium this heating is balanced mainly by infrared cooling from carbon dioxide, there must be a local temperature maximum, so that heat can radiate to cooler regions.

If radiative control of temperature were to continue down to the ground, the temperature in the troposphere would decrease much more rapidly with height than is observed, and this temperature profile (and its associated density profile) would be statically unstable. Such a temperature profile could not persist, but might be expected to give rise to convective activity that would modify the temperature profile, causing it to decrease less rapidly with height until it was just statically stable again. This process appears to occur in the moist tropical troposphere, where the temperature decrease with altitude is fairly close to the

saturated adiabatic lapse rate, the rate of decrease with height of the temperature of a parcel of air saturated with water vapour, which condenses (releasing latent heat) as the parcel rises; see [Section 2.8](#). At higher latitudes other processes, such as midlatitude cyclones and anticyclones, may also play a part in determining the temperature profile. The position of the tropopause depends on an interplay between the processes that cause the temperature to fall with height in the troposphere and increase with height in the stratosphere; the precise details are a subject of active current research.

The explanations of the cold equatorial tropopause and extremely cold summer mesopause are quite complex. It turns out that there is dynamically driven rising motion in both of these regions; the rising air expands as it moves to lower pressure and cools. Conversely there is descent over the winter pole, leading to temperatures that are warmer than would otherwise be the case.

[Figure 1.6](#) shows the zonal-mean zonal (west-to-east) winds for January. These winds are related to the temperature field in [Figure 1.5](#) by **thermal windshear balance**, as expressed

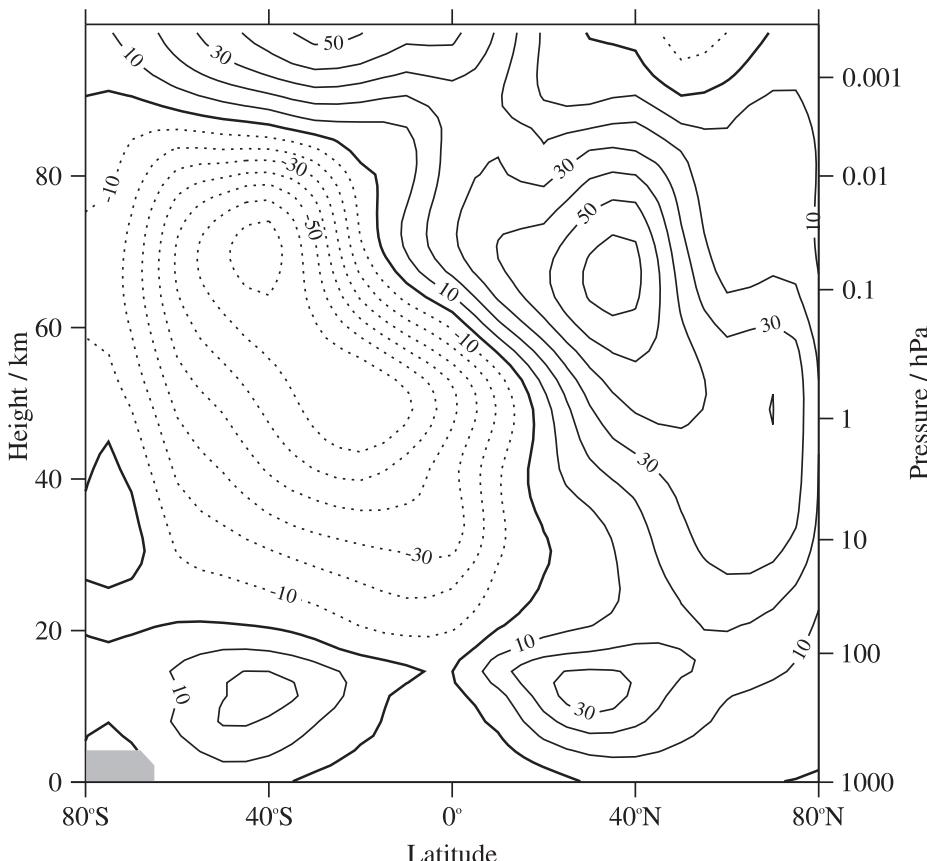


Fig. 1.6

Zonal-mean zonal wind (m s^{-1}) for January, from the CIRA dataset. Thin solid lines: eastward winds; thick solid line: zero winds; dashed lines: westward winds. A small region at low levels over Antarctica is omitted. Based on data from [Fleming et al. \(1990\)](#).

by equation (4.28b). In the troposphere, the mean zonal winds are generally eastward at midlatitudes, with two prominent ‘jet streams’, and westward at low latitudes. (The terms ‘westerly’, meaning eastward, and ‘easterly’, meaning westward, are commonly used in meteorology, but will be avoided in this book.) In the stratosphere and mesosphere the mean zonal winds are generally eastward in winter and westward in summer. The winds at low latitudes shown in Figure 1.6 are not representative of every January, partly because they do not follow a simple annual cycle in the equatorial lower stratosphere; there is a prominent **quasi-biennial oscillation** there, with a period of approximately 28 months.

1.4.2 Gravity waves

Figure 1.7 shows northward and eastward wind components between altitudes of 60 and 80 km, over Alaska, measured by a ground-based radar. The radar transmits radio waves almost vertically and measures the backscattered signal; from this, wind velocities can be determined; see Section 7.3.2. Roughly sinusoidal wind fluctuations in the vertical are seen, with a wavelength of about 15 km and propagating downwards in time with a period of about 9 h.

These quasi-sinusoidal fluctuations are strongly suggestive of some kind of wave motion, and further investigation confirms this. The waves are an example of the fluid-dynamical gravity waves mentioned in Section 1.1. Atmospheric gravity waves are analogous to horizontally propagating surface waves on water, which depend on the restoring mechanism provided by the contrast in density between air and water: the water in a wave crest is denser

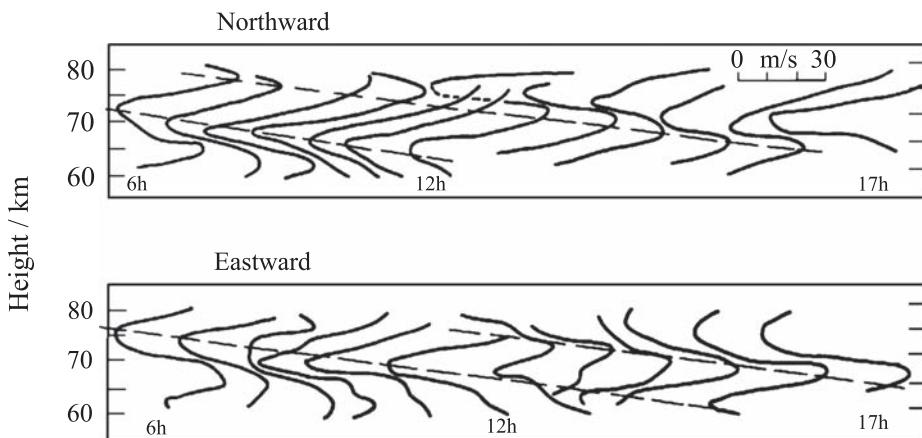


Fig. 1.7

An example of an atmospheric gravity wave over Alaska, as measured by a ground-based radar. The northward and eastward wind components (m s^{-1}) are shown as a function of height, between altitudes of 60 and 80 km, and at each hour over a period of about 17 hours. From Balsley *et al.* (1983).

than the surrounding air and tends to fall, while the air in a wave trough is lighter than the surrounding water and tends to rise. We can imagine the smooth vertical density variation in the atmosphere to be approximated by a stack of thin fluid layers, whose densities decrease with height. The surface-wave mechanism operates at each density interface, so now the wave can propagate vertically as well as horizontally.

The particular type of gravity wave shown in [Figure 1.7](#) is called an inertia–gravity wave (see [Section 5.4](#)); it is actually of large enough horizontal scale (a few hundred kilometres) and period to be influenced to some extent by the Earth’s rotation. These measurements provide an unusually clear example of a sinusoidal oscillation: in most cases, too many other dynamical processes are occurring in the atmosphere for waves to be very easily identified, and careful data analysis must be performed to isolate them.

Perhaps surprisingly, the downward phase progression of the waves in [Figure 1.7](#) indicates upward propagation of ‘information’ by the waves (i.e. an upward group velocity). Gravity waves will be studied in detail in [Section 5.4](#) and, among other things, it will be shown that this type of wave is **dispersive**; the phase and group velocities can therefore be in different directions. This is just one of the ways in which the propagation characteristics of the atmospheric waves studied in this book differ from those of the more familiar non-dispersive waves such as electromagnetic waves in a vacuum.

Gravity waves are generated in many different ways, including by air flow over mountains and by convective activity in the troposphere. Waves generated in the lower atmosphere may propagate upwards into the stratosphere and mesosphere. As the background air density decreases, the amplitudes of the wave fluctuations in wind (and associated fluctuations in temperature and density) will rise. As a result, gravity waves may attain large amplitudes in the mesosphere and exert a considerable influence on the mean atmospheric state there.

1.4.3 Rossby waves

[Figure 1.8](#) depicts the temperature in the Northern Hemisphere, at a level near 24 km altitude, during a period when the stratosphere was disturbed by a vigorous dynamical event known as a **stratospheric warming**. A cold region is located on one side of the pole (roughly along 0° E) and a warm region on the other (roughly along 180° E). Moving around a latitude circle near the pole, we find that the temperature varies roughly sinusoidally with longitude, one wavelength encompassing 360° of longitude. The fact that the cold part of the disturbance is smaller and stronger than the warm part shows that the fluctuation is not exactly sinusoidal. However, the phenomenon is wave-like in many respects and is an example of the Rossby wave mentioned in [Section 1.1](#). The horizontal wavelength is several thousand kilometres in extent and the wave’s dynamics are quite different from those of the gravity wave. The propagation mechanism is rather subtle, depending both on the rotation and on the curvature of the Earth; the details are given in [Section 5.5](#). Rossby waves, like gravity waves, are dispersive.

Several types of Rossby wave are observed in the atmosphere. Some are **stationary**, that is, their wave patterns are fixed with respect to the Earth; since they are dispersive

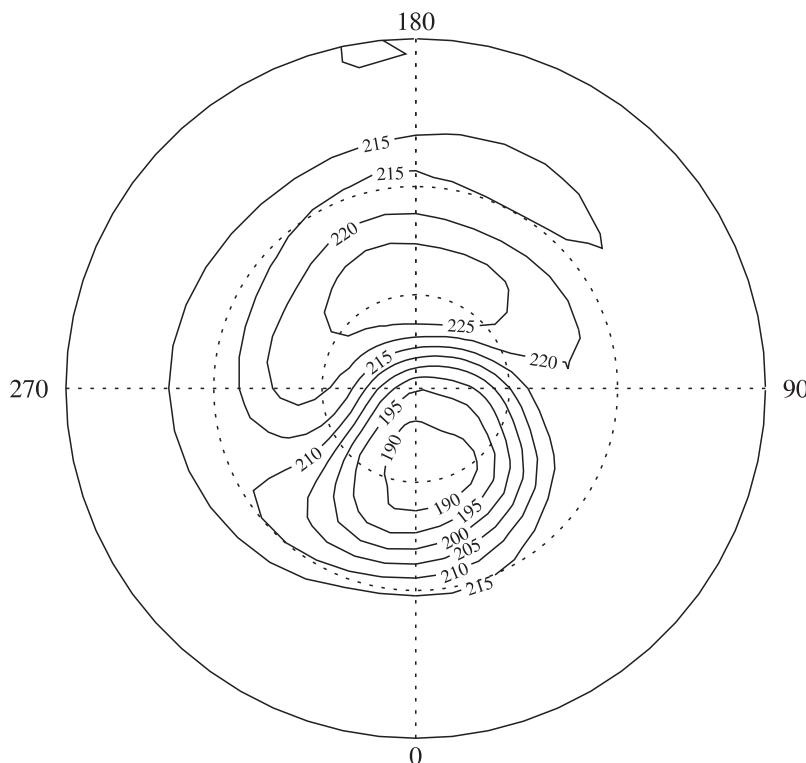


Fig. 1.8 Polar stereographic map of atmospheric temperature (K) near 24 km altitude in the Northern Hemisphere stratosphere on 9 January 1992, as measured by the Improved Stratospheric and Mesospheric Sounder (ISAMS) on the Upper Atmospheric Research Satellite (UARS). The North Pole is at the centre, the 60° N and 30° N latitude circles are shown and the equator is the outer circle; four longitudes are also shown. (Diagram supplied by Dr A. M. Iwi.)

they may still propagate information, because the group velocity may be non-zero even if the phase speed is zero. Others have patterns that travel with respect to the Earth. Strong disturbances in the stratosphere, such as that shown in Figure 1.8, are frequently due to upward propagation of Rossby waves generated by large-scale weather disturbances in the troposphere. It is found that, when the background winds are eastward, as in winter (see Figure 1.6), only the longest-wavelength stationary Rossby waves can propagate vertically. This accounts for the observed prevalence of larger-scale disturbances in the winter stratosphere.

1.4.4 Ozone

As mentioned in Section 1.4.1, atmospheric ozone is important because it absorbs solar ultra-violet radiation, thus protecting human and animal life from potentially harmful

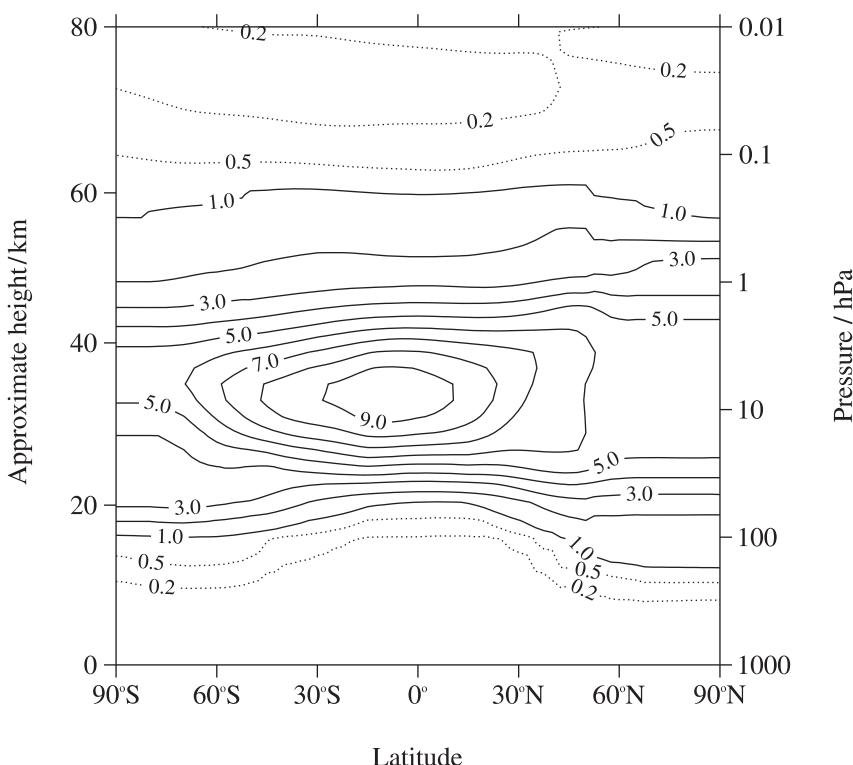


Fig. 1.9

Zonal-mean volume mixing ratio of ozone (parts per million by volume), as a function of latitude and height, for January, based on the 5-year climatology of Li and Shine (1995). Data provided by Dr D. Li.

consequences. It is therefore crucial to have good measurements of ozone concentrations and a good understanding of the processes by which it is produced and destroyed.

A typical vertical profile of the ozone number density was shown in Figure 1.4. An alternative measure of the concentration of a gaseous constituent of air, such as ozone, is the **volume mixing ratio**, that is, the number density of the constituent divided by the total number density of the ‘air’. A useful property of the volume mixing ratio is that, unlike the number density, it is constant for a moving parcel of air in the absence of chemical production and loss processes. Figure 1.9 gives a latitude–height cross-section of the zonal-mean volume mixing ratio of ozone for January. This shows that the maximum values are in the low-latitude stratosphere (where the photochemical production of ozone is greatest) but also that there are significant values over the winter pole, where no production takes place since the Sun is below the horizon all day. This suggests that ozone is transported into the ‘polar night’ region by wind motions.

Further information on the global ozone distribution is given in Figure 1.10, which shows the ‘column ozone’, a measure of the total number of ozone molecules in a vertical column of atmosphere, as a function of latitude and season. Low values of column ozone are found all year round at low latitudes, despite the fact that this is where most production of ozone takes place. Maximum amounts are found in spring: in the Northern Hemisphere

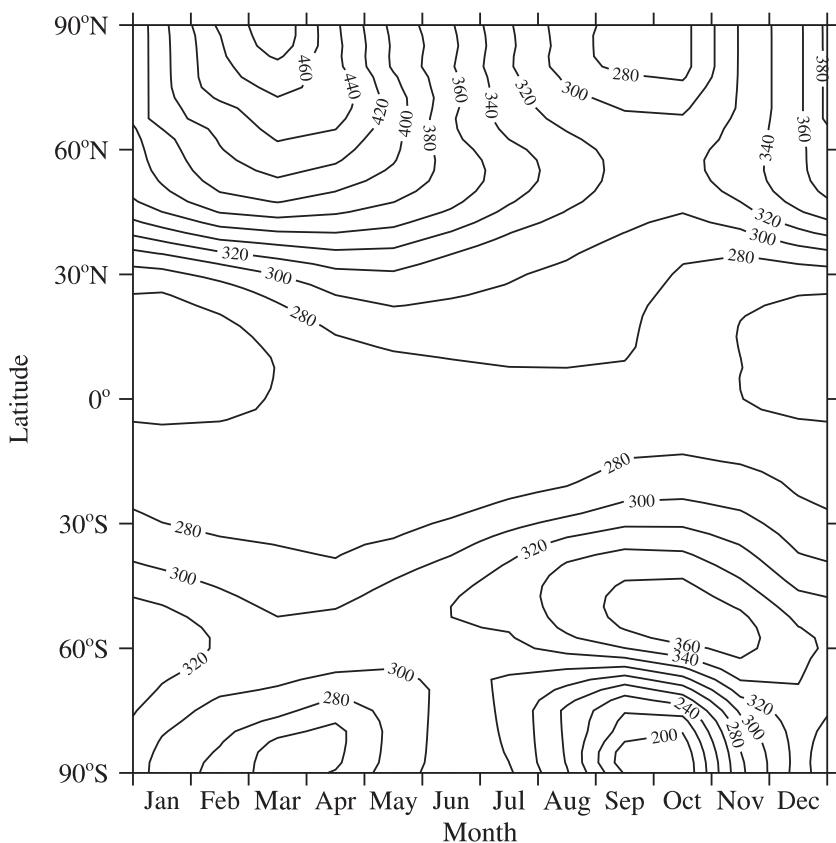


Fig. 1.10 The observed annual cycle in column ozone, based on the 5-year climatology of Li and Shine (1995). The units are Dobson Units: see Section 6.7. Data provided by Dr D. Li.

this maximum occurs at high latitudes, whereas in the Southern Hemisphere it occurs at middle latitudes; a relative minimum is found in the Antarctic spring.

The Antarctic spring minimum of column ozone in Figure 1.10 is a manifestation of the **Antarctic ozone hole**, a dramatic reduction of ozone in the Antarctic stratosphere that has been observed in spring since the mid-1970s. Intensive international efforts, both observational and theoretical, established that this loss of ozone is due to complex chemical and physical processes, involving chlorine resulting from man-made compounds such as chlorofluorocarbons (CFCs), taking place on the ice and water particles that can form in the extremely low Antarctic winter temperatures. The ozone lost during spring over Antarctica is mostly replenished there later in the year, but this effect may be contributing to a slow global loss of ozone. This research led to the Montreal Protocol, in which a phasing out of CFCs was agreed by the international community. Depletion of ozone is also occurring at other latitudes and is currently the subject of global monitoring and modelling. Further details are given in Chapter 6.

1.5 Weather and climate

The word **weather**, while having a clear enough meaning to the layperson, does not have a precise definition in atmospheric physics. However, it tends to encompass tropospheric events associated with atmospheric flows with length scales of hundreds of metres and more, and time scales of a few days or less, although weather patterns may occasionally persist for a week or two. Weather phenomena are notoriously irregular: indeed, Lorenz's famous pioneering work on chaotic systems originated in a study of a simple nonlinear atmospheric model and was motivated by considerations of weather prediction. However, atmospheric data averaged over a month or so usually behave in a more regular manner. Most localities have regular longer-term variations in the average weather conditions, with a roughly annual cycle. This annual cycle does not repeat precisely from one year to the next: there may be dramatic interannual variations in the average weather at a given place.

Much of the thrust of atmospheric research has been directed towards the improvement of weather forecasting, for periods of up to two weeks ahead. Although present-day forecasting is partly limited by the insufficiency of computer power and incompleteness of atmospheric data, it is also limited to some extent by gaps in our understanding of the basic physics and dynamics of the atmosphere.

The word **climate** refers to the state of the atmosphere on longer time scales, typically averaged over several years or more. The understanding of climate and climate change does not necessarily require a complete understanding of every weather event; conversely there are physical processes, operating on long time scales, that are unimportant for weather prediction but crucial for climate prediction. (An example is heat transport from the deep ocean, which may vary on decadal or longer time scales.)

There is much current concern about whether human activity may be changing the climate significantly, principally through an enhancement of the greenhouse effect due to increasing levels of carbon dioxide resulting from the burning of fossil fuels and the destruction of the tropical rain forests. Given the long time scales involved, the detection of changes in climate is difficult; however, the latest estimates suggest, for example, that the global mean surface air temperature has increased by about $0.7 \pm 0.2^\circ\text{C}$ over the last century. It is even more difficult to distinguish between man-made and natural climate change, given the complexity of the climate system (including the atmosphere, oceans, ice cover, biosphere and solid Earth). Since controlled experiments cannot be performed on the climate system, we must use models to identify cause-and-effect relationships, as noted in [Section 1.2](#). Even the most sophisticated current models are still over-simplified. However, the Fourth Assessment Report of the Intergovernmental Panel on Climate Change ([Solomon et al. \(2007\)](#)) states with very high confidence that the global average net effect of human activities since 1750 has been one of warming, amounting to an increase of between 0.6 and 2.4 W m^{-2} in the power per unit area entering the climate system. See [Chapter 8](#) for more details.

Further reading

More advanced treatments of some of the atmospheric physical processes discussed in this chapter are given in the texts by Houghton (2002), Goody (1995) and Salby (1996). The book by Wallace and Hobbs (2006) is an introductory text, with a sound physical basis and with meteorological applications in mind. Holton (2004) provides an excellent introduction to atmospheric dynamics as do Marshall and Plumb (2008) for both the atmosphere and the ocean. A non-technical account of atmospheric waves is given by Andrews (1991). A lucid introduction to the basic physical principles of climate change, intended for non-specialists, is given by Archer (2007). See Lorenz (1993) for a fascinating description of the influence of atmospheric research on the development of chaos theory.

In this chapter we show how basic thermodynamic concepts can be applied to the atmosphere. We first note in [Section 2.1](#) that the atmosphere behaves as an ideal gas. Some basic information on the various gases comprising the atmosphere is presented in [Section 2.2](#). The fact that the atmosphere is fairly close to being in hydrostatic balance is used in [Section 2.3](#) to develop some very simple ideas about the vertical structure of the atmosphere. An important quantity related to entropy, the **potential temperature**, is discussed in [Section 2.4](#). The concept of an **air parcel** is introduced in [Section 2.5](#) and is used to develop ideas about atmospheric stability and buoyancy oscillations. A brief introduction to the concept of **available potential energy** is given in [Section 2.6](#).

The rest of the chapter is devoted to the thermodynamics of **water vapour** in the air. [Section 2.7](#) recalls the basic thermodynamics of phase changes and introduces several measures of atmospheric water vapour content. These ideas are exploited in [Section 2.8](#), in which some effects of the release of latent heat are investigated in a calculation of the **saturated adiabatic lapse rate**, which gives information on the stability of a moist atmosphere. The **tephigram**, a graphical method of representing the vertical structure of temperature and moisture and calculating useful physical results, is introduced in [Section 2.9](#). Finally, some of the basic physics of the formation of cloud droplets by condensation of water vapour is considered in [Section 2.10](#).

2.1 The ideal gas law

To a good approximation the atmosphere behaves as an ideal (or perfect) gas, with each mole of gas obeying the law

$$pV_m = RT, \quad (2.1)$$

where p is the pressure, V_m is the volume of one mole, R is the universal gas constant and T is the absolute temperature. We can obtain the corresponding law for unit mass of air by noting that if the mass of one mole is M_m then the density $\rho = M_m/V_m$. So, from equation (2.1),

$$p = \frac{RT}{V_m} = \frac{R}{M_m} T \rho$$

and hence

$$p = R_a T \rho, \quad (2.2)$$

where $R_a \equiv R/M_m$ is the gas constant per unit mass of air. The value of R_a depends on the precise composition of the sample of air under consideration.¹

2.2 Atmospheric composition

Consider a small sample of air of volume V , temperature T and pressure p , containing a mixture of gases G_i ($i = 1, 2, \dots$). If there are n_i molecules of gas G_i in the sample, then the total number of molecules in the sample is

$$n = \sum n_i, \quad (2.3)$$

where the sum is taken over all the gases in the mixture, and the total mass of the sample is

$$m = \sum m_i n_i, \quad (2.4)$$

where m_i is the molecular mass of gas G_i .

We define the **mass mixing ratio** μ_i of gas G_i as the total mass of the molecules of gas G_i in the sample, divided by the total mass of the complete sample.² Thus

$$\mu_i = \frac{m_i n_i}{m}. \quad (2.5)$$

We now introduce the ideal gas law in the form

$$pV = nk_B T, \quad (2.6)$$

where k_B is Boltzmann's constant. (The connection with the molar form, equation (2.1), can be seen by noting that, for one mole, $n = N_A$, where N_A is Avogadro's number, and recalling that $R = N_A k_B$.) The **partial pressure** p_i of gas G_i is the pressure that would be exerted by the molecules of G_i from the sample if they alone were to occupy volume V at temperature T ; from equation (2.6)

$$p_i = n_i \frac{k_B T}{V}. \quad (2.7)$$

Similarly, the **partial volume** V_i of gas G_i is the volume that would be occupied by the molecules of gas G_i from the sample if they, alone, were to be held at temperature T and pressure p ; again from equation (2.6)

$$V_i = n_i \frac{k_B T}{p}. \quad (2.8)$$

¹ Note that many meteorology texts use R for the gas constant per unit mass of air: however, we follow standard physics practice and use R for the *molar* gas constant.

² When the gas under consideration is water vapour, it is the practice to define the mass mixing ratio as the mass of water vapour divided by the total mass *minus* the mass of water vapour, i.e. by the mass of dry air in the sample. The mass of water vapour divided by the total mass is called the **specific humidity**. However, in most cases when the mass mixing ratio is used it is a small number (e.g. < 0.03 for water vapour and less still for other trace gases such as carbon dioxide and ozone), so the difference between the two definitions is also small, and we shall ignore it in this book.

(Note that Dalton's laws of partial pressures, $p = \sum p_i$, and partial volumes, $V = \sum V_i$, follow immediately from these definitions and equation (2.3).) From equations (2.5)–(2.7) we can relate the mass mixing ratio to the partial pressure as follows:

$$\mu_i = \frac{n m_i p_i}{m p} = \frac{m_i}{\bar{m}} \frac{p_i}{p}, \quad (2.9)$$

where

$$\bar{m} = m/n \quad (2.10)$$

is the mean molecular mass for the sample. We also define the **volume mixing ratio** v_i (also known as the **mole fraction**) by

$$v_i = \frac{V_i}{V};$$

by equations (2.6)–(2.8) we have

$$v_i = \frac{n_i}{n} = \frac{p_i}{p}. \quad (2.11)$$

Note that the two mixing ratios are related by

$$\mu_i = \frac{m_i}{\bar{m}} v_i.$$

Another measure of the concentration of an atmospheric gas is the **number density** (the number of molecules of the gas per unit volume), n_i/V . If we wish to follow the motion of the sample of air, the number density may change either through changes of the volume V of the sample or through changes of n_i resulting from chemical reactions. For many purposes the mass and volume mixing ratios are more convenient measures of concentration when the transport of chemicals is being studied, since they are affected not by volume changes but only by chemical production or loss. For other purposes, partial pressure is sometimes used to quantify chemical concentrations.

The Earth's atmosphere is composed mainly of nitrogen and oxygen, with a much smaller amount of carbon dioxide and still less of certain trace gases such as ozone. See Table 2.1 for a list of some of the more important species. We shall see in Chapters 3, 6 and 8 that some gases such as carbon dioxide, water vapour and ozone are of crucial importance in determining the structure of the atmosphere, despite the fact that they are present only in small amounts. This is because of their ability to absorb and emit infra-red radiation, which is not shared by nitrogen and oxygen.

From equations (2.4), (2.10) and (2.11) the mean molecular mass of an air sample is

$$\bar{m} = \frac{m}{n} = \sum m_i \frac{n_i}{n} = \sum m_i v_i.$$

Similarly, the mean *molar* mass \bar{M} is the mean of the molar masses M_i of the constituent gases G_i , weighted by the volume mixing ratios:

$$\bar{M} = \sum M_i v_i.$$

Using Table 2.1 it can be verified that the mean molar mass of dry air is about 28.97.

Table 2.1 Some gases in the atmosphere. The unit ppmv (parts per million by volume) is used here for CO₂ and O₃; this is a standard unit of volume mixing ratio for minor species. The volume mixing ratios are fairly uniform throughout the lower and middle atmosphere for well-mixed gases that are mostly chemically inert, namely N₂, O₂, CO₂ and Ar. However, the volume mixing ratio for CO₂ is increasing by about 19 ppmv per decade: the value quoted is the global and annual mean for 2009 (see Dr Pieter Tans, NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends/). The unit of molar mass is g mol⁻¹ or equivalently kg kmol⁻¹.

Gas	Volume mixing ratio	Molar mass	Distribution
Nitrogen, N ₂	0.78	28.02	Well-mixed
Oxygen, O ₂	0.21	32.00	Well-mixed
Carbon dioxide, CO ₂	386 ppmv	44.01	Well-mixed
Water vapour, H ₂ O	≤ 0.03	18.02	Maximum in troposphere
Ozone, O ₃	≤ 10 ppmv	48.00	Maximum in stratosphere
Argon, Ar	0.0093	39.95	Well-mixed

2.3 Hydrostatic balance

For an atmosphere at rest, in static equilibrium, the net forces acting on any small portion of air must balance. Consider for example a small cylinder of air, of height Δz and horizontal cross-sectional area ΔA , as depicted in Figure 2.1. This is subject to a gravitational force $g \Delta m$ downwards, where its mass $\Delta m = \rho \Delta A \Delta z$ and g is the gravitational acceleration (assumed constant throughout this book). This force must be balanced by the difference between the upward pressure force $p(z) \Delta A$ on the bottom of the cylinder and the downward pressure force $p(z + \Delta z) \Delta A$ on the top. We therefore have

$$g\rho \Delta A \Delta z = p(z) \Delta A - p(z + \Delta z) \Delta A;$$

by cancelling ΔA and using the Taylor expansion

$$p(z + \Delta z) \approx p(z) + \frac{dp}{dz} \Delta z,$$

we get the equation for **hydrostatic balance**,

$$\frac{dp}{dz} = -g\rho. \quad (2.12)$$

(In Chapter 4 we extend this approach to the case in which the portion of air is accelerating and therefore no longer in static equilibrium.)

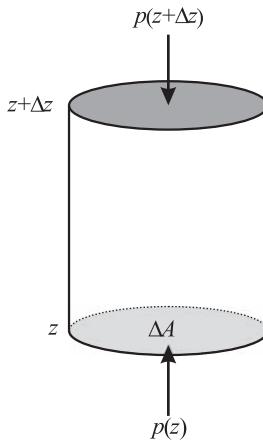


Fig. 2.1

The vertical pressure forces acting on a small cylinder of air.

We can derive some basic properties of the atmosphere, given that it is an ideal gas and assuming that it is in hydrostatic balance. First eliminating the density ρ from equations (2.2) and (2.12), we obtain a useful alternative form of the hydrostatic balance equation,

$$\frac{dp}{dz} = -\frac{gp}{R_a T}. \quad (2.13)$$

If the temperature is a known function of height, $T(z)$, we can in principle find the pressure and density as functions of height also. Equation (2.13) can be rewritten

$$\frac{d}{dz}(\ln p) = -\frac{g}{R_a T}$$

and this may be integrated in z from the ground (say $z = 0$) upwards, given the pressure at the ground (say p_0):

$$\ln p - \ln p_0 = -\frac{g}{R_a} \int_0^z \frac{dz'}{T(z')}$$

or, taking exponentials,

$$p = p_0 \exp\left(-\frac{g}{R_a} \int_0^z \frac{dz'}{T(z')}\right). \quad (2.14)$$

The simplest case is that of an isothermal temperature profile, i.e. $T = T_0 = \text{constant}$, when the pressure decays exponentially with height:

$$p = p_0 \exp\left(-\frac{gz}{R_a T_0}\right) = p_0 e^{-z/H}, \quad (2.15)$$

where $H = R_a T_0 / g$ is the **pressure scale height**, the height over which the pressure falls by a factor of e . In this isothermal case the density also falls exponentially with height in the same way: $\rho = \rho_0 \exp(-z/H)$, ρ_0 being the density at the ground. For an isothermal atmosphere with $T_0 = 260$ K, H is about 7.6 km.

The **lapse rate** Γ denotes the rate of decrease of temperature with height:

$$\Gamma(z) = -\frac{dT}{dz};$$

in general the temperature decreases with height ($\Gamma > 0$) in the troposphere and increases with height ($\Gamma < 0$) in the stratosphere; see Figure 1.3. A layer in which the temperature increases with height ($\Gamma < 0$) is called an **inversion layer**. If Γ is constant in the region between the ground and some height z_1 , say, then the temperature in that region decreases linearly with height and the integral in equation (2.14) can again be evaluated explicitly; see Problem 2.3.

Another useful deduction from the hydrostatic equation in the form (2.13) is the ‘thickness’, or depth, of the layer between two given surfaces of constant pressure. Suppose that the height of the pressure surface $p = p_1$ is z_1 and the height of the pressure surface $p = p_2$ is z_2 . Then, if $p_1 > p_2$, we must have $z_1 < z_2$, since pressure decreases with height when hydrostatic balance applies. From equation (2.13), $g dz = -R_a T d(\ln p)$; integration gives

$$z_2 - z_1 = -\frac{R_a}{g} \int_{p_1}^{p_2} T d(\ln p).$$

The integral can in principle be evaluated if the temperature T is known as a function of pressure p : this may be provided for example by a weather balloon or a satellite-borne instrument. In particular, if T is constant,

$$z_2 - z_1 = \frac{R_a T}{g} \ln\left(\frac{p_1}{p_2}\right);$$

if T is not constant, we can still write

$$z_2 - z_1 = \frac{R_a \bar{T}}{g} \ln\left(\frac{p_1}{p_2}\right),$$

provided that we define \bar{T} as a suitably weighted mean temperature within the layer:

$$\bar{T} = \frac{\int_{p_2}^{p_1} T d(\ln p)}{\int_{p_2}^{p_1} d(\ln p)}.$$

Thus the thickness of the layer between two pressure surfaces is proportional to the mean temperature of that layer.

2.4 Entropy and potential temperature

The First Law of Thermodynamics, applied to a small change to a closed system, such as a mass of air contained in a cylinder with a movable piston at one end (see Figure 2.2), can be written

$$\delta U = \delta Q + \delta W, \quad (2.16)$$

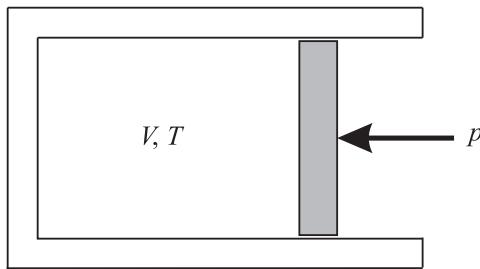


Fig. 2.2 A cylinder of air of volume V , at pressure p and temperature T , closed by a movable piston (shaded).

where δU is the increase of internal energy of the system in the process, δQ is the heat supplied to the system and δW is the work done on the system. In terms of functions of state, equation (2.16) can be written

$$\delta U = T \delta S - p \delta V, \quad (2.17)$$

where S is the **entropy** of the system. An alternative form of equation (2.17) is

$$\delta H = T \delta S + V \delta p, \quad (2.18)$$

where $H = U + pV$ is the **enthalpy**. Since equations (2.17) and (2.18) involve functions of state, they apply both for reversible and for irreversible changes. However, we shall mostly restrict our attention to **reversible** changes, for which the equations

$$\begin{aligned} \delta Q &= T \delta S, \\ \delta W &= -p \delta V \end{aligned} \quad (2.19)$$

also hold.

For *unit mass* of ideal gas, for which $V = 1/\rho$, it can be shown that

$$U = c_v T, \quad (2.20)$$

where c_v is the specific heat capacity at constant volume and is independent of T . Therefore the ideal gas law, equation (2.2), implies that, for unit mass of air,

$$H = c_v T + R_a T = c_p T, \quad (2.21)$$

where $c_p = c_v + R_a$ is the specific heat capacity of air at constant pressure. On substituting the expression (2.21) and $V = 1/\rho = R_a T/p$ into equation (2.18), we get

$$T \delta S = c_p \delta T - \frac{R_a T}{p} \delta p. \quad (2.22)$$

Division by T gives

$$\delta S = c_p \frac{\delta T}{T} - R_a \frac{\delta p}{p} = c_p \delta(\ln T) - R_a \delta(\ln p), \quad (2.23)$$

and integration gives the entropy per unit mass

$$S = c_p \ln T - R_a \ln p + \text{constant} = c_p \ln(Tp^{-\kappa}) + S_0, \quad (2.24)$$

where $\kappa = R_a/c_p$, which is approximately $\frac{7}{5}$ for a diatomic gas, and S_0 is a constant.

An **adiathermal** process is one in which heat is neither gained nor lost, so that $\delta Q = 0$. An **adiabatic** process is one that is both adiathermal and reversible; from equation (2.19) it follows that $\delta S = 0$ for such a process. Imagine a cylinder of air, originally at temperature T and pressure p , that is compressed adiabatically until its pressure equals p_0 . We can find its resulting temperature, θ say, using equation (2.23) together with the fact that $\delta S = 0$ for an adiabatic process, so that

$$c_p \delta(\ln T) = R_a \delta(\ln p).$$

Integrating and using the end conditions $T = \theta$ and $p = p_0$ then gives

$$c_p \ln\left(\frac{\theta}{T}\right) = R_a \ln\left(\frac{p_0}{p}\right),$$

and hence, using $\kappa = R_a/c_p$ again,

$$\theta = T\left(\frac{p_0}{p}\right)^\kappa. \quad (2.25)$$

The quantity θ is called the **potential temperature** of a mass of air at temperature T and pressure p . The value of p_0 is usually taken to be 1000 hPa. Using equation (2.24) it follows that the potential temperature is related to the specific entropy S by

$$S = c_p \ln \theta + S_1,$$

where S_1 is another constant. By definition, the potential temperature of a mass of air is constant when the mass is subject to an adiabatic change; conversely, the potential temperature will change when the mass is subject to a non-adiabatic (or **diabatic**) change. As we shall see, the potential temperature is often a very useful concept in atmospheric thermodynamics and dynamics.

2.5 Parcel concepts

We have just discussed adiabatic processes for a mass of air contained in a cylinder. To apply similar concepts to the atmosphere, we introduce the idea of an **air parcel** – a small mass of air that is imagined to be ‘marked’ in some way, so that its passage through the surrounding air (‘the environment’) can in principle be traced. The parcel is influenced by the environment, but we assume that it does not itself change the environment. The pressure within the parcel is taken to equal that of the surrounding environment, but its temperature, density and composition may differ from those of the environment. The parcel concept is useful, but should not be taken too literally; for example, a real mass of air will rapidly mix with its surroundings and will also inevitably influence the surrounding air.

One simple way to think of an air parcel is to imagine it to be enclosed in a thin balloon of negligible surface tension and heat capacity. We may also take the balloon to have negligible thermal conductivity, in which case the parcel moves adiabatically if there are no sources or sinks of heat within it.³ In the adiabatic case, we can extend the definition of the potential temperature from a cylinder of air to a parcel; it is the final temperature θ of a parcel that is imagined to be brought adiabatically from pressure p and temperature T to pressure p_0 .

For an adiabatically rising parcel, the potential temperature and entropy are constant as its height changes, so we can write

$$\left(\frac{d\theta}{dz}\right)_{\text{parcel}} = 0, \quad \left(\frac{dS}{dz}\right)_{\text{parcel}} = 0.$$

From equation (2.23) we therefore have the following relation between the vertical derivatives of temperature and pressure, following the parcel:

$$0 = \frac{c_p}{T} \left(\frac{dT}{dz}\right)_{\text{parcel}} - \frac{R_a}{p} \left(\frac{dp}{dz}\right)_{\text{parcel}},$$

so that

$$-\left(\frac{dT}{dz}\right)_{\text{parcel}} = -\frac{R_a T}{c_p p} \left(\frac{dp}{dz}\right)_{\text{parcel}} = \frac{g}{c_p} \equiv \Gamma_a, \quad (2.26)$$

say, where equations (2.12) and (2.1) have been used. The quantity Γ_a is the rate of decrease of temperature with height, following the adiabatic parcel as it rises. It is called the **adiabatic lapse rate**; when applied to a mass of dry air, it is called the **dry adiabatic lapse rate** (DALR) and is approximately 9.8 K km^{-1} .

An alternative derivation of the expression (2.26) for the DALR is to note that, for unit mass undergoing a reversible change,

$$\delta Q = T \delta S = c_p \delta T - \frac{R_a T}{p} \delta p = c_p \delta T - \frac{\delta p}{\rho} = c_p \delta T + g \delta z, \quad (2.27)$$

from equations (2.19) and (2.22), the ideal gas law (2.2) and the hydrostatic equation (2.12). For adiabatic motion of the parcel $\delta Q = 0$ and so, letting $\delta z \rightarrow 0$,

$$-\frac{dT}{dz} = \frac{g}{c_p} = \Gamma_a,$$

as before.

The actual lapse rate $-dT/dz$ in the atmosphere will generally differ from the DALR. To investigate the implications of this, consider a parcel that is originally at equilibrium at height z , with temperature T , pressure p and density ρ , all equal to the values for the surroundings. Now suppose that an instantaneous upward force is applied to the parcel, so that it rises adiabatically through a small height δz , without influencing its surroundings; see Figure 2.3.

³ Such heat sources could be due, for example, to latent heating or cooling; see Section 2.7.

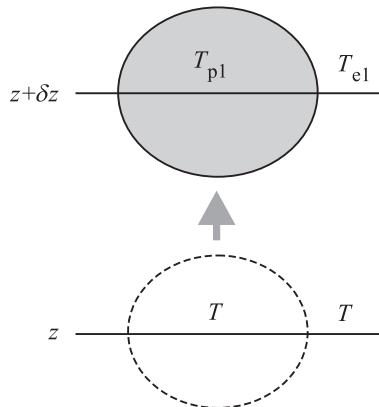


Fig. 2.3 A parcel (shown shaded) displaced a height δz from its equilibrium position at height z (shown dashed).

At the displaced position $z_1 = z + \delta z$ the parcel temperature has increased to T_{p1} , say, according to the adiabatic lapse rate:

$$T_{p1} = T + \left(\frac{dT}{dz} \right)_{\text{parcel}} \delta z = T - \Gamma_a \delta z. \quad (2.28)$$

On the other hand, the environment temperature at height z_1 is

$$T_{e1} = T + \left(\frac{dT}{dz} \right)_{\text{env}} \delta z = T - \Gamma \delta z. \quad (2.29)$$

If $\Gamma \neq \Gamma_a$ there is therefore a temperature difference between the displaced parcel and its surroundings.

However, since the pressures are the same inside and outside the parcel at height z_1 , these pressures are both equal to

$$p_1 = p + \left(\frac{dp}{dz} \right)_{\text{env}} \delta z.$$

By the ideal gas law, equation (2.2), the densities inside and outside the parcel are

$$\rho_{p1} = \frac{p_1}{R_a T_{p1}}, \quad \rho_{e1} = \frac{p_1}{R_a T_{e1}},$$

respectively. The volume of the parcel at height z_1 equals the volume of air displaced there; therefore, if $\rho_{p1} > \rho_{e1}$, the mass of the parcel at z_1 is greater than the mass of air displaced, so the parcel is ‘heavier’ than its surroundings. This holds provided that the temperature of the parcel is less than that of its surroundings ($T_{p1} < T_{e1}$), which in turn is true if $\Gamma < \Gamma_a$, from equations (2.28) and (2.29), i.e. provided that the environment temperature falls less rapidly with height than the adiabatic lapse rate; see Figure 2.4. In this case the displaced parcel, being denser than its surroundings, will tend to fall back towards its equilibrium level; we say that the atmosphere is **statically stable** (or ‘stable’, for short) near height z .

On the other hand, if $\Gamma > \Gamma_a$, so that the environment temperature falls more rapidly with height than the adiabatic lapse rate, a parcel displaced adiabatically upwards finds

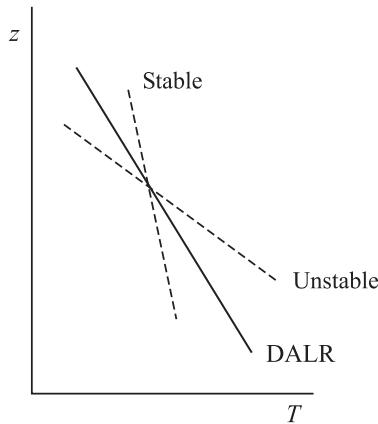


Fig. 2.4 Schematic height profiles of temperature, showing the dry adiabatic lapse rate (DALR) and examples of stable and unstable environmental lapse rates.

itself ‘lighter’ than its surroundings and hence continues to rise: the atmosphere is then said to be **statically unstable** (or ‘unstable’) near z ; see Figure 2.4 again. If $\Gamma = \Gamma_a$ we have neutral stability.

As a further step, we can investigate the buoyancy force on the parcel and its acceleration. The upward buoyancy force on the parcel is g times the difference between the mass of air displaced and the mass of the parcel, i.e.

$$gV_1(\rho_{e1} - \rho_{p1})$$

at height z_1 , where V_1 is the volume of the parcel there. By Newton’s second law, this force can be equated to the mass of the parcel times its acceleration, i.e.

$$\rho_{p1} V_1 \frac{d^2(\delta z)}{dt^2}.$$

We therefore have

$$\begin{aligned} \frac{d^2(\delta z)}{dt^2} &= g \left(\frac{\rho_{e1}}{\rho_{p1}} - 1 \right) = g \left(\frac{T_{p1}}{T_{e1}} - 1 \right) \\ &= g \left(\frac{T - \Gamma_a \delta z}{T - \Gamma \delta z} - 1 \right) = g \left(\frac{\Gamma - \Gamma_a}{T - \Gamma \delta z} \right) \delta z \\ &= -\frac{g}{T} (\Gamma_a - \Gamma) \delta z \end{aligned}$$

to leading order in the small quantity δz . (The ideal gas law (2.2) has been used to go from the second to the third expression here.) We therefore have an equation of the form

$$\frac{d^2(\delta z)}{dt^2} + N^2(\delta z) = 0, \quad (2.30)$$

where

$$N^2 = \frac{g}{T} (\Gamma_a - \Gamma) = \frac{g}{T} \left(\frac{dT}{dz} + \frac{g}{c_p} \right), \quad (2.31)$$

and the temperature T in equation (2.31) is that of the environment.

In the case of a statically stable region of the atmosphere, in which $\Gamma_a > \Gamma$, equation (2.31) indicates that $N^2 > 0$ and equation (2.30) represents simple harmonic motion, with sinusoidal solutions. The parcel then oscillates up and down at an angular frequency N , where N is called the **buoyancy frequency** or the **Brunt–Väisälä frequency**; for the lower atmosphere the corresponding period $2\pi/N$ is a few minutes (see Problem 2.7). For a statically unstable region of the atmosphere, in which $\Gamma_a < \Gamma$, $N^2 < 0$ so that N is imaginary, which leads to exponential solutions of equation (2.30), one of which corresponds to the displaced parcel continuing to move at an increasing speed.

The quantity N^2 is a useful measure of atmospheric stratification. It can be related to the potential temperature of the environment as follows. Taking logarithms of equation (2.25) and differentiating, we obtain

$$\begin{aligned} \frac{1}{\theta} \frac{d\theta}{dz} &= \frac{1}{T} \frac{dT}{dz} - \frac{\kappa}{p} \frac{dp}{dz} = \frac{1}{T} \frac{dT}{dz} + \frac{\kappa \rho g}{p} \\ &= \frac{1}{T} \frac{dT}{dz} + \frac{g}{c_p T}, \end{aligned} \quad (2.32)$$

using the ideal gas law (2.2) and the relation $c_p = R_a/\kappa$. On combining equations (2.31) and (2.32) we obtain the equation

$$N^2 = \frac{g}{\theta} \frac{d\theta}{dz}.$$

Thus a region of the atmosphere is statically stable if θ increases with height and is statically unstable if θ decreases with height.⁴

2.6 The available potential energy

It is frequently useful to consider the energy content of the atmosphere, so some basic concepts are presented here. Consider first the potential energy E_P of a vertical column of atmosphere, of unit horizontal cross-section, stretching from the ground (where $z = z_0$ and $p = p_0$, say) to a great height (where $z = \infty$ and $p = 0$). Since the potential energy of a small slice, of thickness Δz , of this column is $(\rho \Delta z)gz$, the potential energy of the whole column is

$$E_P = \int_{z_0}^{\infty} \rho g z dz.$$

⁴ Note that, for a compressible atmosphere, the static stability depends on the vertical increase of the potential temperature, not on the vertical decrease of the density, as might have been expected intuitively. The latter is applicable only to an incompressible fluid. See Problem 2.8.

However, $\rho g = -dp/dz$ from the hydrostatic equation (2.12), so we can write

$$E_P = - \int_{z_0}^{\infty} z \frac{dp}{dz} dz = \int_{p=0}^{p=p_0} z dp,$$

where the integration variable has been changed from z to p and the upper and lower limits of integration have been reversed, with a corresponding change of sign. We can integrate the last expression in this equation by parts to get

$$E_P = [zp]_{p=0}^{p=p_0} + \int_{z_0}^{\infty} p dz.$$

Now assuming that $zp \rightarrow 0$ as $z \rightarrow \infty$ and $p \rightarrow 0$,⁵ we get

$$E_P = \int_{z_0}^{\infty} p dz + E_0, \quad \text{where} \quad E_0 = z_0 p_0. \quad (2.33)$$

The boundary term $E_0 = 0$ if the surface elevation $z_0 = 0$, but in general we must allow for topography of varying height over the Earth's surface, with $z_0 \neq 0$.

We now consider the internal energy of the air in the column. As noted in equation (2.20) the internal energy per unit mass is $c_v T$, so the integral of this over the mass of the column is

$$E_I = \int_{z_0}^{\infty} \rho c_v T dz = \frac{c_v}{R_a} \int_{z_0}^{\infty} p dz, \quad (2.34)$$

where the ideal gas law (2.2) has been used. Note that E_I is proportional to the integral in equation (2.33). We define the **total potential energy** E_T of the column as the sum of the potential and internal energies so, from equations (2.33) and (2.34),

$$E_T = E_P + E_I = \frac{c_v + R_a}{R_a} \int_{z_0}^{\infty} p dz + E_0 = \frac{1}{\kappa} \int_{z_0}^{\infty} p dz + E_0,$$

since $c_v + R_a = c_p = R_a/\kappa$. Note that the integral here is the total enthalpy of the column, since the enthalpy per unit mass is $c_p T$, from equation (2.21), so that the enthalpy per unit volume is $c_p \rho T = c_p p / R_a$, using the ideal gas law, equation (2.2).

We can extend these ideas to the atmosphere as a whole by integrating the column values over the Earth's surface area; for example, the global total energy is

$$\langle E_T \rangle = \iint_{\text{surface}} E_T dx dy,$$

where x and y are horizontal coordinates.

Given any atmospheric state, it is of interest to ask whether a suitable redistribution of mass can lead to a state of lower total potential energy, thus releasing energy that can be converted, say, to kinetic energy of motion. The answer in general is yes, but it turns out that the maximum amount of potential energy that can be released is in practice very much less

⁵ This holds in all reasonable cases; for example, p decays exponentially with z in an isothermal atmosphere, as shown in equation (2.15).

than $\langle E_T \rangle$. Lorenz showed how the state of minimum total potential energy (or **reference state**) can be calculated, assuming an *adiabatic* redistribution of the atmospheric mass.

A crucial part of Lorenz's calculation was to note that, given an adiabatic redistribution of mass, the mass of air above any given surface of constant potential temperature θ (an **isentropic surface** or **isentrope**) will not change. This is because, in an adiabatic process, each small mass element of air must retain its potential temperature and must therefore remain on a given isentropic surface. Consider then the mass M_1 above the isentrope $\theta = \theta_1$, say, in the actual atmospheric state; this is

$$M_1 = \iint_{\text{surface}} dx dy \int_{z_1}^{\infty} \rho dz,$$

where $z_1(x, y)$ is the height of the isentrope. From the hydrostatic equation (2.12) this equals

$$M_1 = \frac{1}{g} \iint_{\text{surface}} dx dy \int_0^{p_1} dp = \frac{1}{g} \iint_{\text{surface}} p_1(x, y) dx dy = \frac{1}{g} \langle p_1 \rangle,$$

where $p_1(x, y)$ is the pressure on the isentrope.⁶ In particular the total mass of the atmosphere is

$$M_s = \frac{1}{g} \iint_{\text{surface}} p_0(x, y) dx dy = \frac{1}{g} \langle p_0 \rangle.$$

The same results must hold for the reference state so that, under an adiabatic redistribution of mass, the global integral $\langle p_1 \rangle$ (or global mean) of the pressure on the θ_1 isentrope (and therefore on every isentrope) must remain unchanged.

Together with other arguments, this allows the reference state and also the difference in total potential energy between the actual and reference states to be calculated. This difference is called the **available potential energy** and represents the maximum amount of potential energy that might be released for conversion into, say, kinetic energy. The details of the calculation are generally quite complicated; however, a simple special case is considered in [Problem 2.9](#).

2.7 Moisture in the atmosphere

It was mentioned in [Section 2.2](#) that water vapour is a minor constituent of the atmosphere. However, despite its low and variable volume mixing ratio ($\lesssim 0.03$), water vapour plays a crucial role in weather and climate processes. It is responsible for the precipitation (rain and snow) that forms such an important part of our weather; it has significant radiative effects (see [Chapter 3](#)); and, through polar stratospheric clouds, it is associated, in its ice

⁶ This is the global integral of the result that, under hydrostatic balance, the pressure at a given level equals g times the mass per unit horizontal area of air above that level: see also [Problem 2.1](#).

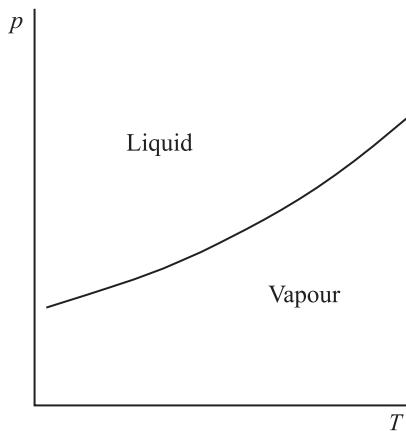


Fig. 2.5 The phase transition between liquid and vapour, portrayed in a schematic temperature–pressure diagram. The curve indicates the temperature variation of the saturation vapour pressure.

and liquid phases, with important chemical reactions that lead to the depletion of polar ozone (see Chapter 6).

In the remainder of this chapter we consider the precipitation properties of water vapour; these result from the physical fact that, at terrestrial atmospheric temperatures and pressures, water undergoes **phase changes** between the vapour, liquid and solid states.

We recall some basic thermodynamics of phase changes, focusing first on the vapour–liquid phase transition at a plane interface. At this transition, the vapour is *saturated*,⁷ that is, in equilibrium with the liquid. The rates of evaporation and condensation are equal, with as many molecules escaping from unit area of the liquid surface per unit time as return. The phase transition is conveniently portrayed in a p – T diagram (Figure 2.5), which indicates the temperature variation of the pressure at which the phase transition takes place (the **equilibrium vapour pressure** or the **saturation vapour pressure**, SVP). The slope of this curve is given by the **Clausius–Clapeyron equation**

$$\frac{dp}{dT} = \frac{\delta S}{\delta V} = \frac{L}{T \delta V}; \quad (2.35)$$

where δS is the entropy gained as unit mass of water changes from liquid to vapour (it is positive, since the vapour exhibits greater disorder than does the liquid state), δV is the increase of volume as unit mass changes from liquid to vapour (it is also positive) and L is the latent heat of vaporization per unit mass (also called the specific enthalpy of vaporization), given by $L = T \delta S$. In fact the specific volume V_v of the water vapour is much greater than that of the liquid, so $\delta V \approx V_v = 1/\rho_v = R_v T/p$, using the ideal gas law, equation (2.2), for the vapour; here ρ_v is the vapour density and R_v is the specific gas constant for the vapour. Hence the Clausius–Clapeyron equation can be written in the more convenient form

⁷ See Bohren (1987) for a clear exposition of the misleading nature of the adjective *saturated* in this context: in particular for its suggestion that air can ‘only hold’ a certain amount of water vapour.

$$\frac{dp}{dT} = \frac{Lp}{R_v T^2}, \quad (2.36)$$

where p and T refer to the values at the phase transition.

These considerations apply to water vapour on its own, but can also be applied to water vapour in the presence of ‘air’, provided that p is replaced by the partial pressure of water vapour, traditionally denoted by the symbol e . Using also the notation $e_s(T)$ to denote the saturation vapour pressure (at the phase transition), we can write the Clausius–Clapeyron equation (2.36) in the form

$$\frac{de_s}{dT} = \frac{Le_s}{R_v T^2}. \quad (2.37)$$

Note that if L is constant (a fairly good approximation at typical atmospheric temperatures), this can be integrated to give

$$e_s(T) = e_s(T_0) \exp \left(\frac{L}{R_v} \left(\frac{1}{T_0} - \frac{1}{T} \right) \right), \quad (2.38)$$

where T_0 is a constant reference temperature.

We can relate the partial pressure e of water vapour to its volume mixing ratio ν and mass mixing ratio μ , say, using equations (2.9) and (2.11):

$$\nu = \frac{e}{p}, \quad \mu = \epsilon \frac{e}{p}, \quad (2.39)$$

where p is the total air pressure, as usual, and

$$\epsilon = \frac{m_v}{\bar{m}} \approx \frac{m_v}{m_d} = \frac{18.02}{28.97} = 0.622.$$

Here m_v is the molecular mass of water vapour, \bar{m} is the mean molecular mass of moist air and m_d is the molecular mass of dry air (i.e. the constituents of the air other than water vapour). (Molar masses may also be used here.) Since the mixing ratio of the water vapour is so small ($\mu < 3 \times 10^{-2}$), $\bar{m} \approx m_d$. Another important measure of the moisture content of air is the **relative humidity**, defined by

$$RH = \frac{e}{e_s(T)}$$

and usually expressed as a percentage.

Now consider a parcel of moist air, of unit mass, containing mass μ of water vapour, so that μ is the mass mixing ratio of water vapour⁸ according to our definition (2.5). So long as no condensation or evaporation takes place, this mass μ remains constant. The vapour pressure of the parcel satisfies

$$e = \frac{\mu p}{\epsilon} \quad (2.40)$$

by equation (2.39) and while this remains less than $e_s(T)$ saturation does not occur, since the water vapour in the parcel remains below the vapour pressure curve in Figure 2.5.

⁸ Or, more correctly the specific humidity: see the footnote on page 20. However, as noted there, we shall ignore the small difference between these two quantities.

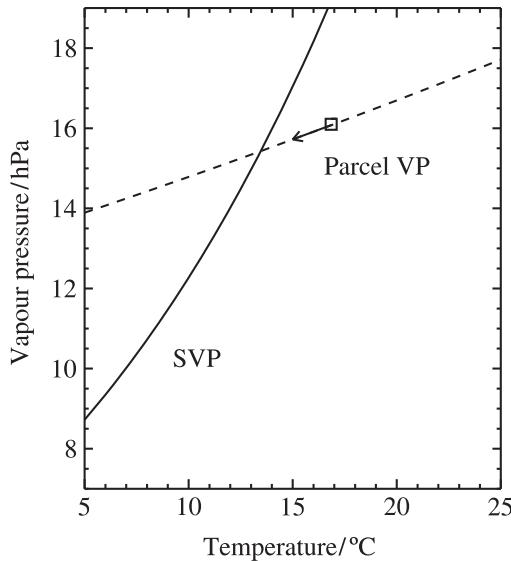


Fig. 2.6 Saturation vapour pressure (SVP) curve (solid) and vapour pressure curve for a parcel of air containing water vapour of mass mixing ratio $\epsilon = 10^{-2} = 10 \text{ g kg}^{-1}$ (dashed). The parcel is taken to start at the point marked by a square, at which $p_0 = 1000 \text{ hPa}$ and $T_0 = 290 \text{ K} \approx 17^\circ\text{C}$. As the parcel rises, its vapour pressure moves down the dashed curve in the direction of the arrow, intersecting the SVP curve at $T = 13.5^\circ\text{C}$ and $e_s = 15.4 \text{ hPa}$. In this case the lifting condensation level is found to be about 960 hPa.

Suppose that the parcel rises adiabatically from the surface (at pressure p_0 and temperature T_0); then the potential temperature of the parcel remains constant at $\theta = T_0$, while the temperature T of the parcel falls according to

$$T = T_0 \left(\frac{p}{p_0} \right)^\kappa \quad (2.41)$$

by equation (2.25). From equations (2.40) and (2.41) we can eliminate the pressure p of the parcel to find how the vapour pressure varies as a function of temperature T , following the motion of the parcel:

$$e_{\text{parcel}}(T) = \frac{\mu p_0}{\epsilon} \left(\frac{T}{T_0} \right)^{1/\kappa}.$$

As illustrated in Figure 2.6, eventually the temperature of the rising parcel falls enough for e_{parcel} to equal e_s and saturation occurs. The pressure level at which this happens can be calculated from equation (2.41): it is called the **lifting condensation level**. (However, it should be noted that in practice liquid need not form at saturation; the vapour may become supersaturated. It is usually necessary for small ‘condensation nuclei’ to be present before liquid drops appear; see Section 2.10.)

A useful related concept is the **saturation mixing ratio**, defined as

$$\mu_s(T, p) = \frac{e_s(T)\epsilon}{p} \quad (2.42)$$

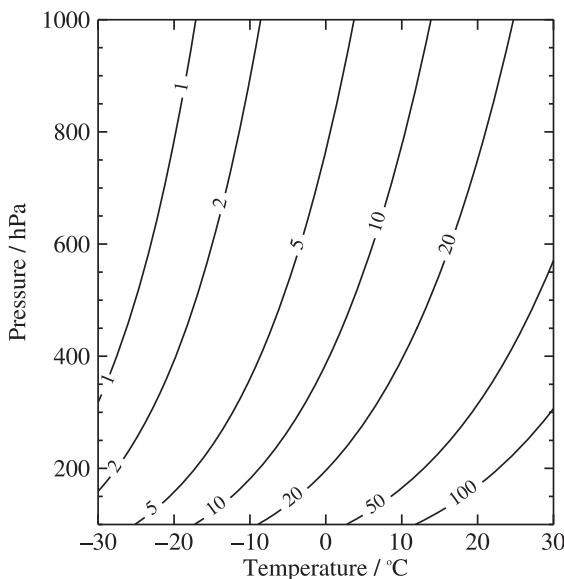


Fig. 2.7

Contours of the saturation mixing ratio $\mu_s(T, p)$, in units of g kg^{-1} .

and conveniently measured in units of grammes per kilogramme (g kg^{-1}). If, at temperature T and pressure p , the mixing ratio μ is less than $\mu_s(T, p)$, then $e < e_s$ from equations (2.40) and (2.42) and the air is unsaturated; it is saturated if $\mu = \mu_s(T, p)$ and is supersaturated if $\mu > \mu_s(T, p)$. Figure 2.7 gives a plot of μ_s as a function of temperature and pressure. The overall behaviour of μ_s is clearly consistent with the facts that it is inversely proportional to pressure and proportional to e_s , which increases with temperature (see Figure 2.6).

The **dew point** T_d of a sample of air is the temperature to which the air must be cooled *at constant pressure* (i.e. *not* following a rising parcel), retaining its water vapour content, for it to become saturated. Therefore, if the water vapour mixing ratio is μ , the dew point T_d satisfies the implicit equation

$$\mu_s(T_d, p) = \mu . \quad (2.43)$$

Equivalently, if the air sample initially has vapour pressure e , then $e_s(T_d) = e$.

We now briefly mention the ice phase. In addition to the vapour–water phase transition just considered, there are also ice–water and ice–vapour transitions, as shown in Figure 2.8. The three transition lines meet at the **triple point**, at the temperature $T_t = 273\text{ K}$ and pressure $p_t = 6.1\text{ hPa}$. The densities and, therefore, the specific volumes V of the three phases are different at the triple point: the specific volume of liquid water is $V_l = 1.00 \times 10^{-3}\text{ m}^3\text{ kg}^{-1}$, the specific volume of ice is $V_i = 1.09V_l$ (unlike most substances, water expands on freezing) and the specific volume of water vapour is $V_v \approx 2 \times 10^5 V_l$.

The basic form (2.35) of the Clausius–Clapeyron equation applies to each of these transitions. Since $V_v \gg V_i$ and the latent heat of sublimation (ice–vapour) is approximately constant, an approximate form of the sublimation curve can be found as in equation (2.38); indeed, since the two latent heats are fairly similar in size, the two curves are quite close.

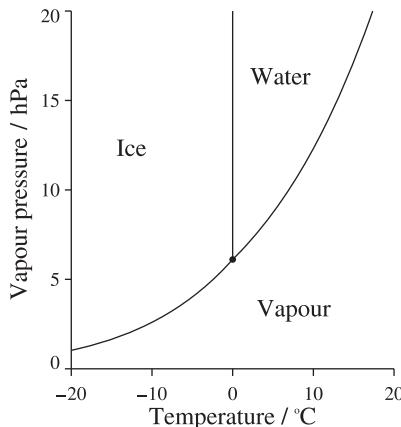


Fig. 2.8

The phase transitions between ice, liquid water and water vapour. The triple point is indicated by the small solid circle. Note that the ice–water curve is not quite vertical, but has a large negative slope of about $-1.4 \times 10^5 \text{ hPa K}^{-1}$ near the triple point.

However, the form of the ice–water transition curve is very different: here δV is small and negative, since V_l is slightly less than V_i , but the latent heat of fusion is positive and non-negligible. Hence the ice–water transition has a large negative slope, as indicated in Figure 2.8.

Many of the concepts developed above for condensation of water vapour at the vapour–liquid transition can be applied also to sublimation at the vapour–ice transition. These include the saturation vapour pressure, saturation mixing ratio and the **frost point**: the temperature to which moist air must be cooled at constant pressure for sublimation to occur.

2.8 The saturated adiabatic lapse rate

While the air in a rising parcel remains unsaturated, the derivation of the adiabatic lapse rate (see equation (2.26)) remains unchanged, apart from the use of the specific heat capacity c_p for the mixture of dry air and water vapour: this is always close to that for dry air alone (see Problem 2.6). However, once saturation takes place, the calculation of the lapse rate following the parcel must be changed significantly, because of the latent heat released.

We extend the second derivation of Section 2.5, considering a saturated parcel (which is taken for convenience to be of unit mass) that rises a distance δz . At saturation, the mixing ratio μ equals the saturation mixing ratio $\mu_s(T, p)$. If a mass $|\delta\mu_s| = -\delta\mu_s$ of water then condenses⁹ during the rise through height δz , an amount of latent heat

$$\delta Q = L |\delta\mu_s| = -L \delta\mu_s \quad (2.44)$$

⁹ Note that $\delta\mu_s$, as is usual for a small change, is defined as an increase in μ_s ; therefore $-\delta\mu_s$ is a decrease of μ_s .

is given to the parcel. The liquid water is assumed to fall out of the parcel and take no further part in its heat balance: this is an irreversible process and it also implies a loss of heat from the parcel. Hence the parcel undergoes a non-adiabatic (and indeed a non-adiathermal) change. However, the amount of heat removed from the parcel by the liquid water is small compared with that remaining in the parcel, so the process is referred to as **pseudo-adiabatic**.

The latent heat release δQ , given by equation (2.44), is equal to the heat input into the parcel while it rises a distance δz and its temperature increases by δT . We assume that this heat input occurs reversibly so

$$\delta Q = c_p \delta T + g \delta z \quad (2.45)$$

by equation (2.27), where c_p is the value for the dry air–water vapour mixture.

On equating the two expressions for δQ in equations (2.45) and (2.44) we obtain

$$c_p \delta T + g \delta z + L \delta \mu_s = 0. \quad (2.46)$$

We now need to express $\delta \mu_s$ in terms of δT and δz . From equation (2.42) we have $\mu_s = e_s/p$. Taking logarithms and differentiating gives

$$\frac{\delta \mu_s}{\mu_s} = \frac{\delta e_s}{e_s} - \frac{\delta p}{p}.$$

However, e_s depends only on T , so $\delta e_s = (de_s/dT) \delta T$; moreover, from the Clausius–Clapeyron equation (2.37),

$$\frac{1}{e_s} \frac{de_s}{dT} = \frac{L}{R_v T^2}.$$

From the hydrostatic equation in the form (2.13) we have $\delta p = -pg \delta z/(R_a T)$, where p is the total pressure; by collecting these results we therefore get

$$\frac{\delta \mu_s}{\mu_s} = \frac{L \delta T}{R_v T^2} + \frac{g \delta z}{R_a T}. \quad (2.47)$$

By eliminating $\delta \mu_s$ from equations (2.46) and (2.47), we obtain

$$\left(c_p + \frac{L^2 \mu_s}{R_v T^2} \right) \delta T + g \left(1 + \frac{L \mu_s}{R_a T} \right) \delta z = 0.$$

Letting $\delta z \rightarrow 0$, we get the **saturated adiabatic lapse rate** (SALR) Γ_s :

$$\Gamma_s = -\frac{dT}{dz} = \frac{g}{c_p} \frac{\left(1 + \frac{L \mu_s}{R_a T} \right)}{\left(1 + \frac{L^2 \mu_s}{c_p R_v T^2} \right)}. \quad (2.48)$$

Note that the factor g/c_p on the right-hand side of equation (2.48) equals the DALR Γ_a .

For typical atmospheric values of T and μ_s it is found that $\Gamma_s \leq \Gamma_a$. Because of the latent heat given to the air by condensation of the water vapour, the temperature drops off less rapidly with height (by about $6\text{--}9 \text{ K km}^{-1}$) at the SALR than it does at the DALR ($\sim 9.8 \text{ K km}^{-1}$, as noted in Section 2.5); see Problem 2.11. Note that Γ_s depends on

the temperature and pressure, through its dependences on T and $\mu_s(T,p)$. Working in terms of the pressure of the parcel, rather than its height, we may show (again using $g\delta z = -R_a T \delta p/p$) that, following the ascending parcel,

$$\frac{dT}{dp} = \frac{\Gamma_s R_a T}{gp} = \Gamma'_s(T,p), \quad (2.49)$$

say. Curves in the T,p plane whose slopes at each point are given by equation (2.49) are called **saturated adiabatics**. Given the expression for Γ'_s and suitable starting values of T and p , they may readily be calculated numerically.

The arguments of Section 2.5 relating to the static stability of a region of the atmosphere can now be extended to allow for the effects of moisture. In particular, it follows that, if the actual lapse rate Γ is less than the SALR Γ_s , then the region is statically stable even if the air is saturated. However, if $\Gamma > \Gamma_s$, a saturated parcel will be unstable. Moreover, if $\Gamma_s < \Gamma < \Gamma_a$, a saturated parcel is unstable but an unsaturated one is not: this situation is called **conditional instability**. See Figure 2.9 for a graphical representation of the various cases considered here: this should be contrasted with the much simpler Figure 2.4 for a dry atmosphere.

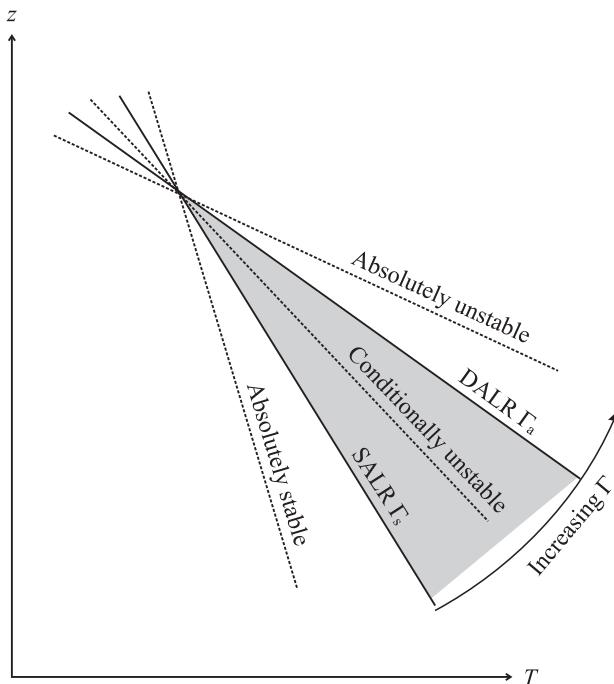


Fig. 2.9

An illustration of how the static stability of a region of the atmosphere varies with different values of the lapse rate $\Gamma = -dT/dz$ when the effects of condensation are included. The solid lines represent the dry adiabatic lapse rate Γ_a and the saturated adiabatic lapse rate Γ_s . The dashed lines show three values of the lapse rate Γ : as Γ increases, the region changes from a state of absolute stability, through a state of conditional instability (in the shaded wedge between Γ_s and Γ_a), to a state of absolute instability.

Under a certain approximation, the saturated adiabatics can be calculated explicitly. First, note that, for reversible processes, $T \delta S = \delta Q$, so equation (2.23) can be written

$$c_p \delta(\ln T) - R_a \delta(\ln p) = \frac{\delta Q}{T} = -\frac{L \delta \mu_s}{T}. \quad (2.50)$$

Provided that the expression on the extreme right-hand side of equation (2.50) can be approximated by $-\delta(L\mu_s/T)$,¹⁰ then

$$\delta \left(c_p \ln T - R_a \ln p + \frac{L\mu_s}{T} \right) = 0.$$

On integrating, dividing by c_p , using $\kappa = R_a/c_p$ and taking exponentials, we get

$$\theta_e(T, p) \equiv T \left(\frac{p}{p_0} \right)^{-\kappa} \exp \left(\frac{L\mu_s}{c_p T} \right) = \text{constant}. \quad (2.51)$$

The quantity θ_e is called the **equivalent potential temperature**. By comparison of equations (2.25) and (2.51) we see that

$$\theta_e(T, p) = \theta(T, p) \exp \left(\frac{L\mu_s(T, p)}{c_p T} \right).$$

Under the given approximation, we have therefore integrated equation (2.49) explicitly, so the curves of constant θ_e closely approximate the saturated adiabatics. It may be shown that, as we follow a saturated adiabatic $\theta_e = \theta_0$, say, to low pressure (and low temperature), it approaches the dry adiabatic $\theta = \theta_0$.

2.9 The tephigram

Meteorologists find it convenient to represent the vertical profiles of atmospheric temperature and moisture on **thermodynamic diagrams**. These profiles may be measured for example by an ascending instrumented balloon (a radiosonde). Thermodynamic diagrams are particularly useful for examining the effects of moisture, for which there are no simple formulae allowing easy analytical calculations.

One such diagram is the **tephigram** shown in Figure 2.10. This uses the temperature T and entropy per unit mass S as orthogonal coordinates.¹¹ The lines of constant S are labelled with the corresponding values of the potential temperature θ (recall that $S = c_p \ln \theta$). They are called **dry adiabatics**; the lines of constant T are called **isotherms**. Since

$$p = p_0 \left(\frac{T}{\theta} \right)^{1/\kappa},$$

¹⁰ This may be shown to hold if $L\mu_s/(c_p T) \ll 1$; since $L/(c_p T) \lesssim 10$ for typical lower-atmospheric temperatures, this approximation holds when $\mu_s \ll 100 \text{ g kg}^{-1}$. From Figure 2.7 it can be seen that this is true over a wide range of temperatures and pressures. See also Problem 2.11.

¹¹ In older works, ϕ was used instead of S , hence the name $T-\phi$ diagram, or tephigram.

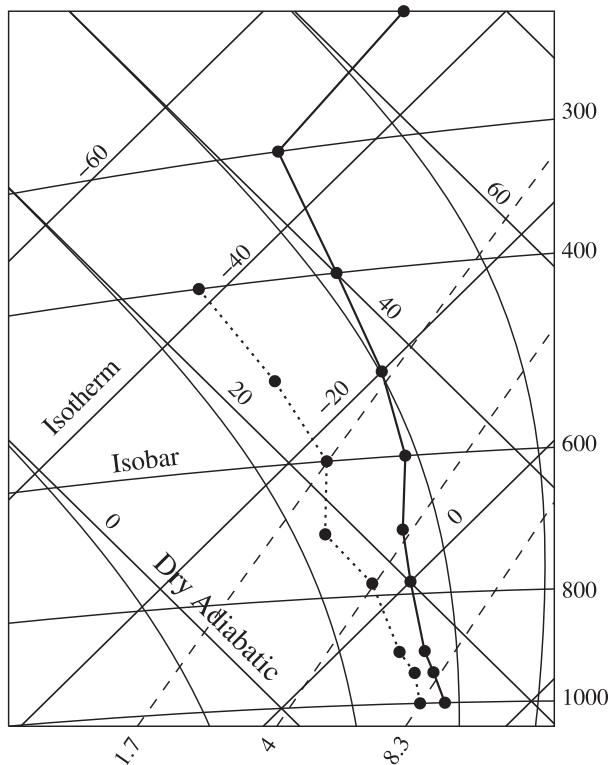


Fig. 2.10 Tephigram, showing isotherms (solid lines sloping up to the right, labelled in $^{\circ}\text{C}$), dry adiabatics (solid lines sloping up to the left, labelled in $^{\circ}\text{C}$) and isobars (slightly curved, roughly horizontal, lines labelled in hPa), lines of constant saturation mixing ratio (dashed, labelled in g kg^{-1}) and saturated adiabatics (strongly curved lines that approach the dry adiabatics near the top of the diagram). Also plotted is the environment curve (heavy solid line segments) and the dew point curve (dotted line segments) for one particular radiosonde ascent.

curves of constant pressure can also be plotted; in the range of T and θ relevant for the lower atmosphere, these are almost straight. The isotherms (parallel to the S axis) and the dry adiabatics (parallel to the T axis) are chosen to point at 45° above and below the horizontal, respectively, so that the curves of constant p (**isobars**) become roughly horizontal.

Also included in Figure 2.10 are two other sets of curves, related to moisture. The first are the lines of constant saturation mixing ratio $\mu_s(T, p)$: these are almost straight and are drawn dashed. (Note that we can if we wish use T and p as independent variables for plotting points on the tephigram, instead of T and S .) They can be plotted using equation (2.42) if an accurate expression for $e_s(T)$ is known. Also plotted are the saturated adiabatics; these are noticeably curved. Each saturated adiabatic can be labelled by the temperature at which it cuts the $p = 1000 \text{ hPa}$ surface (the **wet-bulb potential temperature** θ_w) or the potential temperature which it approaches at low p (the **equivalent potential temperature** θ_e).

During the ascent of a radiosonde the temperature T is measured at a series of pressure levels p , and these can be plotted on the tephigram, to give the **environment curve**. Straight-line segments are drawn between each point, rather than a smooth curve. This gives a representation of a vertical column of atmosphere – but note that this may be slightly misleading, since the balloon takes some time to ascend and also blows some distance downwind as it does so. The dew point is also plotted at each pressure level, giving a separate curve. [Equation \(2.43\)](#) then allows us to find the mixing ratio μ at any pressure level on the environment curve, given the corresponding dew point: we just read off the value of μ_s at the same pressure on the dew point curve.

Many useful results can be obtained from the environment and dew point curves, including inferences about the onset of instability and the formation of clouds: see [Problems 2.12](#) and [2.13](#).

2.10 Cloud formation

In the previous sections we have considered only the case of a flat interface between liquid and vapour. However, this is not directly relevant for the formation of cloud droplets, which are approximately spherical. We shall show in this section that it is essential to allow for the effects of surface tension on cloud droplets. A further finding is that large supersaturation (a vapour pressure e significantly greater than the SVP e_s , or a relative humidity significantly greater than 100%) is required if cloud droplets are to form spontaneously from water vapour: in fact small particles (**cloud condensation nuclei**) are usually needed in the formation process.

For including the effects of surface tension, we introduce the **Gibbs free energy**

$$G = U - TS + pV$$

and note that, using [equation \(2.16\)](#),

$$\delta G = -S \delta T + V \delta p. \quad (2.52)$$

Consider a water droplet immersed in water vapour, at partial pressure e and temperature T (the remaining ‘dry air’ plays no role and can be ignored). Suppose that the liquid and vapour are not in equilibrium, so that $e \neq e_s(T)$, the SVP.¹² Suppose that the Gibbs free energy *per unit mass* of the vapour is $G_v(T, e)$ and that of the liquid is $G_l(T, e)$. Now let the partial pressure be varied slightly from e to $e + \delta e$, while the temperature is held constant. From [equation \(2.52\)](#) it follows that there are small changes to G_v and G_l , given by

$$\delta G_v = V_v \delta e, \quad \delta G_l = V_l \delta e,$$

¹² It can be shown that the difference in pressure between the droplet and the surrounding vapour, due to surface tension, can be ignored here; see [Salby \(1996\)](#), Section 9.2.1.

where V_v and V_l are the specific volumes of the vapour and liquid, respectively. However, $V_v \gg V_l$, hence

$$\delta(G_v - G_l) = (V_v - V_l) \delta e \approx V_v \delta e.$$

The vapour satisfies the ideal gas law $V_v = R_v T/e$, so

$$\delta(G_v - G_l) = R_v T \frac{\delta e}{e} = R_v T \delta(\ln e).$$

Integrating at fixed T gives

$$G_v(T, e) - G_l(T, e) = R_v T \ln e + F(T),$$

where $F(T)$ is a function of integration. However, at equilibrium, on the vapour pressure curve where $e = e_s(T)$, the specific Gibbs free energies G_v and G_l are equal. (This result is used in the standard derivation of the Clausius–Clapeyron equation.) Using this condition to fix $F(T)$, we get

$$G_v(T, e) - G_l(T, e) = R_v T \ln \left(\frac{e}{e_s(T)} \right). \quad (2.53)$$

Now suppose that at some initial time we have a mass m_0 of water vapour, at partial pressure e and temperature T , with no droplet present. The total Gibbs free energy at this time is

$$\mathcal{G}_0 = G_v(T, e)m_0. \quad (2.54)$$

A droplet then starts to condense, at fixed temperature and pressure; suppose that at some later instant its radius is a , so that its surface area is $A = 4\pi a^2$ and its mass is $m_l = 4\pi a^3 \rho_l / 3$ (where ρ_l is the density of the liquid) and the mass of the surrounding vapour is m_v . The total Gibbs free energy of the system is now the sum of the Gibbs free energies of the liquid and vapour, plus a contribution due to surface tension:

$$\mathcal{G} = G_v(T, e)m_v + G_l(T, e)m_l + \gamma A, \quad (2.55)$$

where γ is the surface tension (or the surface energy per unit area). By conservation of mass $m_v = m_0 - m_l$, so, using equations (2.54) and (2.55), we get

$$\mathcal{G} - \mathcal{G}_0 = (G_l - G_v)m_l + \gamma A$$

and, using equation (2.53),

$$\mathcal{G} - \mathcal{G}_0 = -\frac{4}{3}\pi a^3 \rho_l R_v T \ln \left(\frac{e}{e_s(T)} \right) + 4\pi a^2 \gamma. \quad (2.56)$$

The variation with radius a of the total Gibbs free energy of the system therefore takes the form

$$\mathcal{G}(a) = \mathcal{G}_0 - \beta a^3 + \alpha a^2.$$

Figure 2.11 plots this function for two values of the relative humidity e/e_s . Several useful facts can be learned from this figure. Note first that, if $e \leq e_s(T)$ (indicating subsaturated or exactly saturated conditions), then the logarithm in equation (2.56) is negative or zero, so that $\beta \leq 0$ and the curve of $\mathcal{G}(r)$ has no turning point other than $a = 0$. However, if

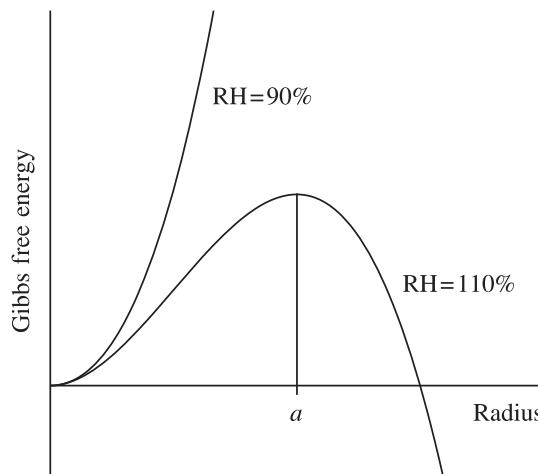


Fig. 2.11 A schematic plot of the Gibbs free energy \mathcal{G} as a function of the droplet radius, for $\text{RH} = e/e_s = 90\%$ and 110% . The second curve reaches a maximum at the radius a given by equation (2.57).

$e > e_s(T)$ (indicating supersaturated conditions), then the logarithm is positive, $\beta > 0$, and there is a maximum of \mathcal{G} at a radius a given by

$$a = \frac{2\alpha}{3\beta} = \frac{2\gamma}{\rho_l R_v T \ln(e/e_s(T))}. \quad (2.57)$$

This is known as **Kelvin's formula**; it may also be written in the form

$$e = e_s(T) \exp\left(\frac{A}{a}\right) \quad \text{where} \quad A = \frac{2\gamma}{\rho_l R_v T}. \quad (2.58)$$

A well-known thermodynamic result is that a system at constant temperature and pressure tends to evolve in such a way that its Gibbs free energy decreases;¹³ stable equilibrium is attained when the Gibbs free energy is a minimum. On the other hand the radius a given by equation (2.57), at which \mathcal{G} is a maximum, corresponds to an unstable equilibrium. Equation (2.58) shows how the partial pressure e of the vapour over a spherical droplet of radius a in this equilibrium state differs from the SVP $e_s(T)$, in the presence of surface tension; see Figure 2.12. Note that an (unstable) equilibrium radius $a = 0.01 \mu\text{m}$ corresponds to a relative humidity $\text{RH} \approx 112\%$, whereas $a = 0.1 \mu\text{m}$ corresponds to $\text{RH} \approx 101\%$. As $a \rightarrow \infty$ we recover the ‘plane surface’ result $\text{RH} \rightarrow 100\%$, i.e. $e \rightarrow e_s$.

The fact that the droplet is in unstable equilibrium at the radius a given by equation (2.57) implies that condensation of a small quantity of vapour increases its radius slightly, and so the droplet continues to grow by further condensation. On the other hand, if a small amount of evaporation occurs at the equilibrium radius, the droplet continues to shrink by evaporation, eventually disappearing altogether.

¹³ See also Section 6.1.

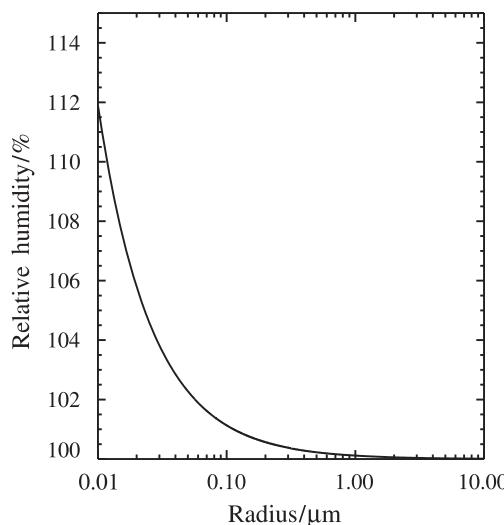


Fig. 2.12 Plot of relative humidity $\text{RH} = e/e_s$ over a spherical droplet in unstable equilibrium, as a function of droplet radius a , at 5 °C. Equivalently, it shows the (unstable) equilibrium radius of a spherical drop as a function of the ambient RH, at 5 °C.

We can therefore see that, if a cloud droplet is to survive, it must somehow attain a radius greater than the equilibrium radius a corresponding to the ambient relative humidity. Given that the RH in clouds is seldom greater than about 101%, Figure 2.12 would require a radius greater than about 0.1 μm. A droplet of this size is unlikely to form by random collisions of smaller droplets, so a more viable process is for the droplet to condense on a small pre-existing solid or liquid particle, known as a **cloud condensation nucleus**.

The above theory must be modified significantly if solutes are present in the cloud droplet. **Raoult's law** states that, over a droplet containing N moles of solute and N_0 moles of pure water, the vapour pressure must be modified by the factor

$$\phi = 1 - \frac{N}{N_0}$$

if $N \ll N_0$. Consider a spherical droplet of radius a and volume $V_a = 4\pi a^3/3$ containing a mass m of solute of molar mass M . Then $N = i_d m/M$, where i_d is the number of ions produced by the dissociation of one solute molecule. For example $i_d = 2$ for NaCl, which dissociates completely into Na^+ and Cl^- ions. Also $N_0 = \rho_l V_a / M_w$ approximately, where ρ_l is the density of the pure water and M_w is the molar mass of pure water. (This approximation ignores the small mass m of solute within the droplet and the fact that the density of the solution is slightly different from ρ_l .) We therefore obtain

$$\phi = 1 - \frac{B}{a^3}, \quad \text{where } B = \frac{3i_d m M_w}{4\pi \rho_l M}; \quad (2.59)$$

combining equations (2.59) and (2.58) we obtain the relative humidity RH, given by

$$\text{RH} \equiv \frac{e}{e_s(T)} = \exp\left(\frac{A}{a}\right) \left(1 - \frac{B}{a^3}\right). \quad (2.60)$$

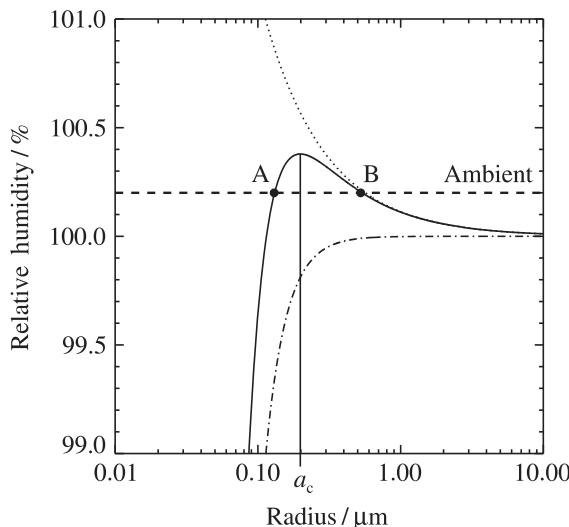


Fig. 2.13

The Köhler curve (solid) for the relative humidity $\text{RH} = e/e_s$ over a spherical droplet of water containing solute, as a function of droplet radius a , at 5°C . The solute is taken to be $10^{-19}\text{ kg of NaCl}$. The Kelvin factor is given by the dotted curve and the Raoult factor is given by the dash-dotted curve. The thick horizontal dashed line and points A and B are discussed in the text.

Figure 2.13 gives an example of this function of radius a : this is called a **Köhler curve**. Also plotted are the Kelvin factor $\exp(A/a)$, which decreases with increasing a , and the Raoult factor $\phi = 1 - (B/a^3)$, which increases with increasing a . The relative humidity has a maximum at $a = a_c$; in this example $a_c \approx 0.2\text{ }\mu\text{m}$, at which value $\text{RH} = \text{RH}_c \approx 100.4\%$.

Consider such a droplet in air with an ambient relative humidity less than this value of RH_c , as given say by the thick dashed horizontal line in Figure 2.13. For this ambient RH there are two equilibrium values of a , on either side of a_c . The smaller value (point A) corresponds to stable equilibrium, for if the RH over the drop falls slightly below the ambient value, corresponding to a slight decrease in the drop's radius, condensation will occur and the radius will increase again to the equilibrium value. Conversely, a slight increase in radius will cause some evaporation, leading to a decrease in radius to the equilibrium value. A stable droplet of this kind is called a **haze** droplet. However, the other equilibrium point, B, will correspond to **instability**: for example, a slight increase in radius will lead to condensation and hence to a further increase in radius. The droplet is then said to be **activated**.

The analysis given above considers only the water vapour in the immediate neighbourhood of the droplet. However, if the droplet is to grow, there must be a continual supply of water vapour to its surface. This can happen by **diffusion** if there is a vapour density gradient in the region surrounding the droplet, with the vapour density increasing with distance. A simple representation of this diffusion is in terms of Fick's law

$$\mathbf{f} = -D\nabla\rho_v,$$

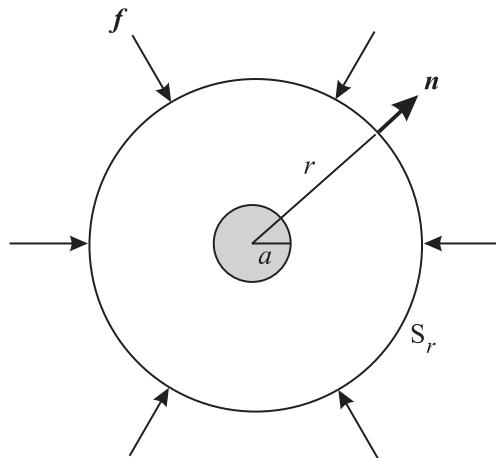


Fig. 2.14 A sketch indicating inward diffusion of water vapour, through a sphere S_r , onto a droplet of radius a .

where ρ_v is the vapour density, f is the vapour-flux vector and D is a diffusion coefficient, assumed constant. Assuming that at some instant the radius of the droplet is a and that the distribution of the vapour density is spherically symmetric, $\rho_v = \rho_v(r)$, the total inward flux of mass of vapour through a sphere S_r of radius $r > a$ is

$$-\int_{S_r} f \cdot n \, dS = 4\pi r^2 \frac{d\rho_v}{dr} D,$$

where n is the outward normal to S_r ; see Figure 2.14.

However, water vapour is lost only by condensation at $r = a$, so for $r > a$ this flux must be independent of r and equal to the rate of increase of mass of the droplet, dM_l/dt . Hence

$$\frac{d\rho_v}{dr} = \frac{dM_l/dt}{4\pi D} \frac{1}{r^2},$$

which can be integrated from $r = a$ to $r = \infty$ to give

$$\rho_v(a) = \rho_v(\infty) - \frac{1}{4\pi Da} \frac{dM_l}{dt}.$$

Therefore, using the ideal gas law $\rho_v = e/(R_v T)$ for the vapour,

$$\frac{dM_l}{dt} = 4\pi Da [\rho_v(\infty) - \rho_v(a)] = \frac{4\pi Da}{R_v} \left[\frac{e(\infty)}{T(\infty)} - \frac{e(a)}{T(a)} \right],$$

where $e(\infty)$ and $T(\infty)$ are the vapour pressure and temperature far from the droplet and $e(a)$ and $T(a)$ are the same quantities at its surface.

Further reading

There are many thermodynamics textbooks covering the material used in this chapter: an excellent, comprehensive modern example is Blundell and Blundell (2009). Good treatments of atmospheric thermodynamics, giving more detail than is presented here, are to be found in Wallace and Hobbs (2006), Salby (1996) and Bohren and Albrecht (1998). The original work on the concept of available potential energy was carried out by Lorenz (1955). More accurate empirical formulae for $e_s(T)$ than equation (2.38) are available: see, for example, Bolton (1980). (One such formula has been used in Figures 2.6, 2.7 and 2.10.) A comprehensive treatment of the tephigram is given by McIntosh and Thom (1983); see also McIlveen (1991). A good text on cloud physics is Rogers and Yau (1989).

Problems

Problem 2.1 Use the hydrostatic equation to show that the mass of a vertical column of air of unit cross-section, extending from the ground to a great height, is p_0/g , where p_0 is the surface pressure. Hence estimate the total mass of the atmosphere. Estimate also the total enthalpy of the atmosphere: this may be taken as a measure of the total heat content of the atmosphere.

In a thought-experiment the temperature of the atmosphere is decreased by 1 K everywhere and the heat thereby released is given to the top 100 m of the ocean, whose temperature rises by an amount δT everywhere. Find δT . Comment on the possible implications of this result for climate change. [Take the specific heat capacity of sea water to be $4.2 \text{ kJ K}^{-1} \text{ kg}^{-1}$.]

Problem 2.2 A balloon is required to carry an instrument payload of 100 kg to an altitude where the pressure is 12 hPa and the temperature is 230 K. The total mass of the payload plus the balloon is 150 kg. What approximate radius of balloon is needed, assuming that after being charged with a measured ‘bubble’ of helium at launch it just becomes spherical at float altitude?

What are the relative merits and disadvantages of balloons, satellites and aircraft as instrument platforms?

Problem 2.3 Show that, if there is a uniform lapse rate Γ , the pressure in the atmosphere is given by

$$p(z) = p_0 \left(1 - \frac{\Gamma z}{T_0}\right)^{g/(\Gamma R_a)}$$

where p_0 is the pressure and T_0 is the temperature at the ground, $z = 0$, and R_a is the gas constant per unit mass of air. Calculate the height at which the pressure is 0.1 of its surface value assuming a surface temperature of 290 K and (i) a uniform lapse rate of 10 K km^{-1} and (ii) a uniform temperature of 290 K.

Problem 2.4 Define the potential temperature θ . What is the temperature as a function of pressure in an atmosphere for which the lapse rate equals the DALR?

Show that an atmosphere of uniform potential temperature must have a finite depth. Calculate this depth for an Earth-like atmosphere with $\theta = 300\text{ K}$.

Problem 2.5 Using an argument similar to that in [Section 2.5](#), but comparing the potential temperature of a parcel with that of the surrounding air, show directly that the air is statically stable if $d\theta/dz > 0$.

Problem 2.6 Derive an expression for the specific heat capacity c_p for air with a water vapour mass mixing ratio of μ . (Use the fact that $c_p = (\partial H/\partial T)_p$, where H is the enthalpy per unit mass; cf. [equation \(2.21\)](#).) Compare the values of the DALR for completely dry air and when $\mu = 0.02$. (The value of c_p for water vapour is given in [Appendix A](#).)

Problem 2.7 Calculate the period of oscillation of an air parcel given that $dT/dz = -6.5\text{ K km}^{-1}$ and $T = 270\text{ K}$.

Problem 2.8 Show that the ‘density lapse rate’ satisfies

$$-\frac{g}{\rho} \frac{d\rho}{dz} = N^2 + \frac{g^2}{c_s^2},$$

where $c_s^2 = pc_p/(\rho c_v)$. ($c_s \approx 330\text{ m s}^{-1}$ is the speed of sound in air.) Comment on the implications for static stability if the density decreases very slowly with height.

Problem 2.9 Estimate the total potential energy per unit area for the Earth’s atmosphere, assuming that it is at a uniform potential temperature of 300 K . An atmosphere similar to that of the Earth is initially all at a uniform potential temperature θ . The air in one hemisphere is heated in such a way that its potential temperature is raised uniformly by an amount $\Delta\theta$. Find an expression for the amount of *available* potential energy per unit area that is generated.

Problem 2.10 Show that the density of water vapour at saturation is given as a function of temperature by

$$\rho_{vs} = \frac{A e^{-T_1/T}}{T},$$

where $T_1 = L/R_v$ and A is a constant. Assuming a constant lapse rate, as in [Problem 2.3](#), show that the following approximations are valid for heights $z \ll T_0/\Gamma$:

$$T^{-1} \approx T_0^{-1} \left(1 + \frac{\Gamma z}{T_0}\right) \approx T_0^{-1} \exp\left(\frac{\Gamma z}{T_0}\right).$$

Hence show that ρ_{vs} decays exponentially with z under these approximations. Evaluate the *e*-folding height (in km) for this decay in a tropical atmosphere with $T_0 = 300\text{ K}$ and $\Gamma = 6.5\text{ K km}^{-1}$ and comment on how it compares with a typical pressure scale height. Check the self-consistency of your approximations in this case.

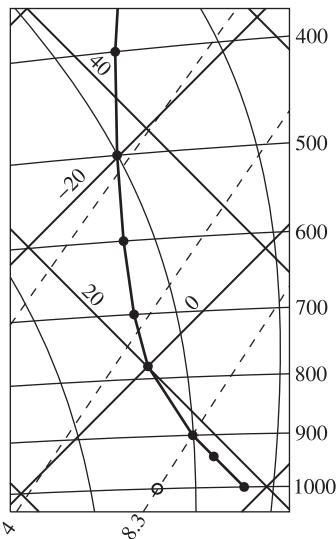


Fig. 2.15

Portion of a tephigram for use in Problem 2.12.

Given a relative humidity of 80%, calculate the total mass of water vapour per unit horizontal area in a vertical column of atmosphere. If all this water vapour were to precipitate out, how many millimetres of rain would result? How much latent heat per unit horizontal area would be released?

Problem 2.11 Estimate the terms $L\mu_s/(R_a T)$ and $L^2\mu_s/(c_p R_v T^2)$ in equation (2.48) for the SALR, given a reasonable estimate of T and p , using Figure 2.7 to estimate μ_s . Hence derive an estimate for the SALR.

Problem 2.12 Figure 2.15 shows a portion of a tephigram. The solid circles represent a temperature sounding and the open circle at 1000 hPa gives the dew point there.

Describe the fate of an air parcel that starts at the ground and rises. At what pressure is its lifting condensation level? What happens to the parcel after condensation starts to occur, and how much further can it rise before it becomes neutrally buoyant with respect to the surrounding atmosphere? Where would you expect clouds to form? (You should include clear statements about the stability or otherwise of the various vertical regions of the atmosphere up to 400 hPa.)

Problem 2.13 Suppose that an imaginary air parcel is moved reversibly around a closed circuit in the tephigram, i.e. that it is forced through a cyclical process in which its temperature, entropy and, therefore, pressure are varied. Using the First Law of Thermodynamics, equation (2.17), show that the work per unit mass done on the parcel is given by the area enclosed by the circuit (described clockwise) in the tephigram. Find the work done on a parcel of mass 1 kg that is moved clockwise around the circuit delineated by the 20 °C and 30 °C dry adiabatics and the 0 °C and -10 °C isotherms.

Problem 2.14 Calculate the times taken for water drops of radii 1, 10 and 100 μm to fall a distance of 1 km in air at the terminal velocity (take the dynamic viscosity η for air to be $1.7 \times 10^{-5} \text{ kg m}^{-1} \text{ s}^{-1}$). Use Stokes' law, which states that the viscous force on a spherical drop of radius r and speed v is $6\pi\eta rv$.

The condition for Stokes' law to hold is that the dimensionless **Reynolds number**, $\text{Re} \equiv \rho vr/\eta$, should be small, where ρ is the air density (see e.g. Acheson (1990)). Check the validity of this condition for each drop radius and comment on your result.

Problem 2.15 Show that the temperature T_i at the surface of a spherical ice crystal of mass M_i growing in a cloud of water droplets at temperature T is given by

$$T_i - T = \frac{L_s}{4\pi\lambda} \frac{\dot{M}_i}{r},$$

where L_s is the specific latent heat of sublimation of ice, \dot{M}_i is the rate of the increase in mass, r is the radius and λ is the thermal conductivity of the air.

Show also that

$$\frac{\dot{M}_i}{r} = \frac{4\pi D}{R_v} \left(\frac{e}{T} - \frac{e_i}{T_i} \right),$$

where D is the diffusion coefficient of water molecules in air, R_v is the gas constant per unit mass of water vapour and e and e_i are the saturation vapour pressures over water and ice respectively.

Given that $(e/T) - (e_i/T_i) = 6.7 \times 10^{-2} \text{ Pa K}^{-1}$ at the relevant temperature (approximately -10°C) calculate (i) the temperature difference between the surface of the ice crystal and the air some distance away and (ii) the time taken for the crystal to grow from 1 to 100 μm radius.

(Take $\lambda = 2.4 \times 10^{-2} \text{ W m}^{-1} \text{ K}^{-1}$ and $D = 0.23 \times 10^{-4} \text{ m}^2 \text{ s}^{-1}$.)

This chapter outlines the basic principles of energy transfer by electromagnetic radiation in the atmosphere. First, in [Section 3.1](#), we introduce the Planck function, the solar spectrum and the concept of local thermodynamic equilibrium. Then in [Section 3.2](#) we list some formal definitions of radiometric quantities and derive and solve the **radiative-transfer equation**, which describes the way in which radiative power is affected by extinction and emission of radiation. In [Section 3.3](#) we present some key facts of molecular spectroscopy and give some of the properties of spectral line shapes. In [Section 3.4](#) we introduce the concept of **transmittance**, the fraction of radiative power that survives propagation from one point to another. In [Section 3.5](#) we apply the concepts introduced in earlier sections to the absorption and emission of infra-red radiation and the absorption of ultra-violet radiation by gases in the atmosphere. This absorption and emission lead to heating and cooling; the principles of the calculation of heating rates are outlined in [Section 3.6](#). In [Section 3.7](#), we revisit the greenhouse effect, investigating two models that are slightly more realistic than that described in [Section 1.3.2](#). Finally, in [Section 3.8](#), we discuss a simple model of atmospheric scattering.

Radiative heating and cooling play crucial roles in the physics of climate change: more details will be given in [Chapter 8](#). The solution of the radiative-transfer equation also underlies certain techniques of atmospheric remote sounding: see [Chapter 7](#).

It is an unfortunate fact that quantitative calculations of radiative heating rates, for example, involve considerable geometric and algebraic detail, which tend to distract attention from the basic physics of the processes. This chapter emphasises the underlying physical concepts as much as possible and keeps the formal mathematics to a minimum; however, the mathematics cannot be avoided entirely.

3.1 Basic physical concepts

The subject of atmospheric radiation is concerned with the transfer of energy within the atmosphere by photons, or equivalently by electromagnetic waves. We recall that the wavelength of visible light lies between the violet at about $0.4\text{ }\mu\text{m}$ (400 nm) and the red at about $0.7\text{ }\mu\text{m}$ (700 nm). Ultra-violet radiation has wavelengths shorter than 400 nm and infra-red radiation has wavelengths longer than 0.7 μm . It is convenient to split the infra-red into the **near infra-red**, between about 0.7 and 4 μm , the **thermal infra-red**, between about 4 and 50 μm , and the **far infra-red**, between about 50 μm and 1 mm.

In the atmosphere, the relevant photons fall naturally into two classes.

- Solar (or short-wave) photons, emitted by the Sun; these correspond to ultra-violet, visible and near infra-red wavelengths between about 0.1 and 4 μm .
- Thermal (or long-wave) photons, emitted by the atmosphere or the Earth's surface; these correspond mainly to thermal infra-red and far infra-red wavelengths, between about 4 and 100 μm .

These two wavelength ranges represent spectral regions of significant black-body emission at temperatures of about 6000 K (a temperature representative of the solar photosphere) and 288 K (the Earth's mean surface temperature), respectively; see [Section 3.1.1](#).

To study the effects of atmospheric radiation, it is necessary to investigate the interaction between photons and atmospheric gases. One way in which solar photons may be lost is by interaction with molecules at certain discrete frequencies, each frequency ν corresponding to an orbital transition of an electron to a higher energy level according to the formula $\Delta E = h\nu$, where ΔE is the difference in energy levels and h is Planck's constant. (The corresponding wavelength λ is given by $\lambda = c/\nu = hc/\Delta E$, where c is the speed of light.) However, the resulting excited state has a limited lifetime and the excitation energy may be lost again in one of two ways.

- (a) The electron falls back to the ground state, re-emitting a photon of the same energy and frequency as the original photon, but in a random direction. This process is called **radiative decay**.
- (b) At sufficiently high pressures, molecular collisions are likely to occur before re-emission takes place, leading to transfer of the excitation energy ΔE to other forms of energy: the photon is then said to have been **absorbed**. If kinetic energy is produced in the process, this will quickly be shared between molecules by collisional interactions and (since thermal energy is the macroscopic expression of molecular kinetic energy) local **heating** of the atmosphere takes place. This transfer of photon energy to heat is called **thermalisation** or **quenching**.

Radiative decay, defined in (a), is an example of the **scattering** of a photon of a given, discrete, frequency by an atmospheric molecule. More important for atmospheric physics, however, is the scattering of photons, over broad ranges of frequencies, by atmospheric molecules and by solid or liquid particles in suspension in the atmosphere. (A suspension of this kind is called an **aerosol**.) In the case of scattering by molecules, whose dimensions are much less than the wavelength of the solar radiation, we have **Rayleigh scattering** (see [Section 7.3.2](#) for more details). In the case of scattering by aerosol particles such as dust and smog, whose dimensions are comparable to the wavelength of the solar radiation, we have the more complex **Mie scattering**. For scattering by particles such as cloud droplets or raindrops, which are much larger than the wavelength of the solar radiation, geometric optics applies; this describes optical phenomena such as rainbows and haloes.

The term **extinction** is used to denote loss of energy from an incoming photon. This can occur either by absorption or by scattering.

Absorption of solar photons may also cause **photo-dissociation**, i.e. the breakdown of the molecules, leading to photochemical reactions, and **photo-ionisation**, in which outer

electrons are stripped from atoms. These interactions occur over a continuous range of frequencies, provided that the energy of the incoming photon is large enough.

Thermal photons may be absorbed and scattered in a similar manner to solar photons. They may also be **emitted**, by the inverse process to absorption, with energy being drawn from molecular kinetic energy, thus leading to a local cooling of the atmosphere. However, at the infra-red frequencies involved here, the relevant ΔE corresponds to a difference between the energies of pairs of vibrationally or rotationally excited states of the emitting molecule, rather than the energy of an electronic transition; see [Section 3.3](#).

In this chapter we shall use the frequency ν and wavelength λ , as appropriate, to describe radiative properties. A related quantity is the wavenumber $\tilde{\nu} = \nu/c = 1/\lambda$; this is commonly used in spectroscopy and is often measured in units of cm^{-1} . It should not be confused with the quantity $k = 2\pi/\lambda$, also called the wavenumber, that is used in the description of wave motion (e.g. [Section 5.4](#)).

3.1.1 The Planck function

An **isothermal cavity** is defined as a cavity whose walls are maintained at a uniform temperature. Under **thermodynamic equilibrium** conditions, the radiation within such a cavity is in equilibrium with the cavity walls and it can be shown that the spectral energy density (the energy per unit volume, per unit frequency interval) depends only on frequency and temperature; the radiation is also isotropic. If a small hole is cut in the cavity then the emitted radiation will have the same form as the radiation within the cavity. Radiation of this kind is called **black-body radiation**.

Planck's law states that the spectral energy density of black-body radiation at absolute temperature T is given by

$$u_\nu(T) = \frac{8\pi h\nu^3}{c^3\{\exp[h\nu/(k_B T)] - 1\}},$$

where k_B is Boltzmann's constant. Since the photons carrying this energy are moving isotropically, the energy density associated with the group of photons moving within a small solid angle $\Delta\Omega$ steradians is $u_\nu \Delta\Omega/(4\pi)$. Consideration of the energy flow per unit time, per unit area, transferred at speed c by this group of photons then shows that the power per unit area, per unit solid angle, per unit frequency interval (the spectral radiance; see [Section 3.2.1](#)) for black-body radiation at temperature T is

$$B_\nu(T) = \frac{2h\nu^3}{c^2\{\exp[h\nu/(k_B T)] - 1\}}; \quad (3.1)$$

this is called the **Planck function**.

The black-body spectral radiance can also be written in terms of the power per unit area, per unit solid angle, per unit **wavelength** interval,

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5\{\exp[hc/(\lambda k_B T)] - 1\}}; \quad (3.2)$$

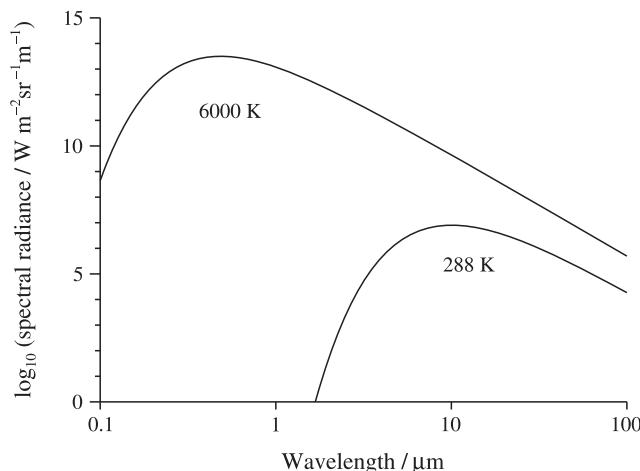


Fig. 3.1 Logarithm of the black-body spectral radiance $B_\lambda(T)$, plotted against the logarithm of wavelength λ , for $T = 6000 \text{ K}$, a typical temperature of the solar photosphere, and 288 K , the Earth's mean surface temperature.

see Problem 3.1. Figure 3.1 shows B_λ for temperatures of 6000 and 288 K. Since large ranges of wavelengths and spectral radiances are under consideration, it is convenient to use a log–log plot here. Note that the curve for 6000 K lies above that for 288 K for all wavelengths: in fact B_λ and B_ν always increase with temperature at fixed λ and ν . (However, when account is taken of the small solid angle subtended by the Sun, the spectral power per unit area reaching a point in the atmosphere from the Sun is much closer in magnitude to that from the Earth: see Problem 3.3.)

If B_λ is integrated over all wavelengths, we obtain the **black-body radiance**

$$\int_0^\infty B_\lambda(T) d\lambda = \frac{\sigma}{\pi} T^4, \quad (3.3)$$

where σ is the Stefan–Boltzmann constant. In terms of an integral over $\ln(\lambda)$, this gives

$$T^{-4} \int_{-\infty}^\infty \lambda B_\lambda(T) d(\ln \lambda) = \frac{\sigma}{\pi}.$$

This suggests plotting $T^{-4}\lambda B_\lambda$ against $\ln \lambda$: the area under the resulting curve is then independent of T . Curves of this kind are shown in Figure 3.2. Note that, with this normalisation, there is little overlap between the black-body spectral radiances at 6000 and 288 K, which are mostly confined to wavelengths shorter and longer than $\lambda = 4 \mu\text{m}$, respectively.

A **black body** is defined as a body that completely absorbs all radiation falling on it. It can be shown that the radiation emitted by a black body is black-body radiation, as defined above. The concept of a black body is an idealisation: a real body will emit less radiation than this. The **spectral emittance** ε_ν of a body is the ratio of the spectral radiance from that body to the spectral radiance from a black body; therefore $\varepsilon_\nu \leq 1$. It follows that a black body emits the maximum possible amount of energy in each frequency interval, at a given

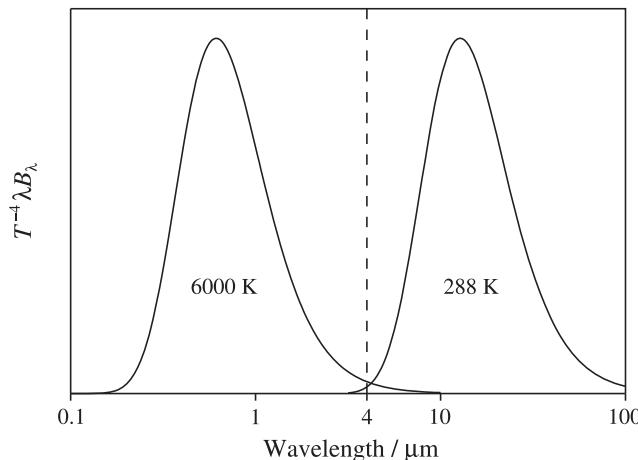


Fig. 3.2 The black-body spectral radiance $B_\lambda(T)$, multiplied by $T^{-4}\lambda$, plotted against the logarithm of wavelength λ , for $T = 6000$ and 288 K. The vertical dashed line at $\lambda = 4\ \mu\text{m}$ roughly separates the solar radiation on the left from the thermal radiation on the right.

temperature. We can also define the **spectral absorptance** α_ν as the fraction of energy per unit frequency interval falling on a body that is absorbed. **Kirchhoff's law** states that $\varepsilon_\nu = \alpha_\nu$; that is, at a given temperature and frequency the spectral emittance of a body equals its spectral absorptance.

3.1.2 Local thermodynamic equilibrium

The standard derivation of the Planck function (3.1) applies to an isothermal cavity containing radiation, but not containing matter. On the other hand, statistical mechanics shows that the energy levels of a material system (for example a gas) in equilibrium at temperature T will be populated according to the **Boltzmann distribution**, when radiation is neglected. That is, the numbers n_1 and n_2 of molecules in states of energy E_1 and E_2 and with statistical weights (or degeneracies) g_1 and g_2 , respectively, are in the ratio given by the Boltzmann distribution

$$\frac{n_1}{n_2} = \frac{g_1}{g_2} \exp\{-(E_1 - E_2)/(k_B T)\}. \quad (3.4)$$

In the case of a gas, this equilibrium ratio is maintained by collisions between the gas molecules.¹

When matter and radiation are both contained in an isothermal cavity and the interaction between the matter and radiation is sufficiently weak, then in thermodynamic equilibrium

¹ For an ideal gas, these collisions must be infrequent; that is, the interaction between the molecules must be weak. However, if equilibrium is to be maintained, some such interaction is essential. The necessity of a weak interaction between different 'aspects' of a system for maintenance of equilibrium is a familiar one in statistical mechanics; see Mandl (1988), page 51.

the radiation will continue to satisfy Planck's law and the matter will continue to satisfy the Boltzmann distribution. The interaction between the matter and the radiation is essential to bring about thermodynamic equilibrium between them, but it must not be so strong as to lead to significant departures of the radiation from Planck's law or of the matter from the Boltzmann distribution.² The interaction will be sufficiently weak, for a given pair of energy levels, if the mean time between collisions for a given molecule, τ_c say (which is inversely proportional to the pressure p), is much shorter than the lifetime for radiative decay, τ_d say, for the given levels. This state of thermodynamic equilibrium will therefore hold if the pressure is large enough, for a given transition between energy states.

The atmosphere does not have a uniform temperature, so we cannot regard it as being in strict thermodynamic equilibrium. However, at high enough pressures, molecular collisions are sufficiently rapid for the Boltzmann distribution (3.4) to hold for each small portion of the atmosphere, given the local value of T . Such a portion of the atmosphere is said to be in **local thermodynamic equilibrium** (LTE) with respect to the given energy states. We can then, for example, use the Planck function (3.1) – which is derived under thermodynamic equilibrium conditions – to represent the spectral radiance.

It can be shown that LTE applies to translational modes (those associated with molecular kinetic energy and macroscopic thermal energy) below about 500 km altitude. For the vibrational and rotational modes involved in the absorption and emission from most radiatively active gases, LTE holds for pressures greater than about 0.1 hPa, corresponding roughly to altitudes below about 60 km. Methods for calculating the spectral radiance when LTE does not hold are complex and will not be discussed in this book.

3.2 The radiative-transfer equation

3.2.1 Radiometric quantities

Several different, but related, quantities are used in the description and measurement of radiation. The most important are as follows.

- The **spectral radiance** (or monochromatic radiance) $L_\nu(\mathbf{r}, \mathbf{s})$ is the power per unit area, per unit solid angle, per unit frequency interval in the neighbourhood of the frequency ν , at a point \mathbf{r} , in the direction of the unit vector \mathbf{s} . It is measured in $\text{W m}^{-2} \text{ steradian}^{-1} \text{ Hz}^{-1}$. The spectral radiance can be visualised in terms of the photons emerging from a small area ΔA with unit normal \mathbf{s} , centred at a point \mathbf{r} : see Figure 3.3. Consider those photons whose momentum vectors lie within a cone of small solid angle $\Delta\Omega$ centred on the direction \mathbf{s} and whose frequencies lie between ν and $\nu + \Delta\nu$. Then $L_\nu \Delta A \Delta\Omega \Delta\nu$ is the energy transferred by these photons, per unit time, from ‘below’ the area ΔA to ‘above’. (Here ‘below’ means in the direction $-\mathbf{s}$ and ‘above’ means in the direction \mathbf{s} .)

² For an advanced treatment, see Landau and Lifshitz (1980), Chapters IV and V.

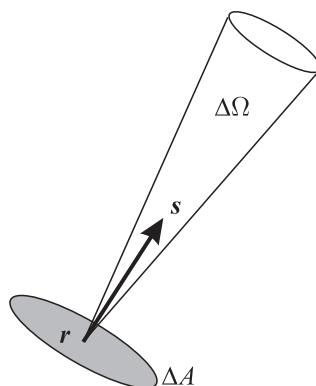


Fig. 3.3

Illustrating the definition of spectral radiance; see the text. Although a cone is shown only emerging from the point r , we must also consider parallel cones emerging from the other points in the area ΔA .

We have already encountered a special case of the spectral radiance, for isotropic black-body radiation in an isothermal cavity (Section 3.1.1), when $L_\nu = B_\nu(T)$, the Planck function; this depends only on the temperature of the cavity and is independent of position and direction.

- The **radiance** $L(\mathbf{r}, \mathbf{s})$ is the power per unit area, per unit solid angle at a point \mathbf{r} in the direction of the unit vector \mathbf{s} ; in other words it is the integral of L_ν over frequency:

$$L(\mathbf{r}, \mathbf{s}) = \int_0^\infty L_\nu(\mathbf{r}, \mathbf{s}) d\nu.$$

Its units are $\text{W m}^{-2} \text{ steradian}^{-1}$.

- The **spectral irradiance** (or monochromatic irradiance) $F_\nu(\mathbf{r}, \mathbf{n})$ is the power per unit area, per unit frequency interval in the neighbourhood of the frequency ν , at a point \mathbf{r} through a surface of normal \mathbf{n} ; its units are $\text{W m}^{-2} \text{ Hz}^{-1}$. It is obtained from the spectral radiance by integration over a hemisphere on one side of the surface:

$$F_\nu(\mathbf{r}, \mathbf{n}) = \int_{2\pi} L_\nu(\mathbf{r}, \mathbf{s}) \mathbf{n} \cdot \mathbf{s} d\Omega(\mathbf{s}), \quad (3.5)$$

where $d\Omega(\mathbf{s})$ is the element of solid angle in the direction \mathbf{s} ; see Figure 3.4. We therefore integrate over all photons in the frequency interval that emerge into the region above the surface.³ As with the Planck function, the spectral radiance and spectral irradiance can alternatively be expressed per unit wavelength interval.

- The **irradiance** (or flux density) $F(\mathbf{r}, \mathbf{n})$ is the power per unit area at a point \mathbf{r} through a surface of normal \mathbf{n} , i.e. the integral of F_ν over frequency, and also the integral of the radiance L over a hemisphere:

³ The angular integration is similar to that used in the kinetic theory of gases; indeed many of the geometric concepts that apply to moving molecules in kinetic theory apply equally to photons, with the simplification that the photons all have the same speed c , rather than a distribution of speeds.

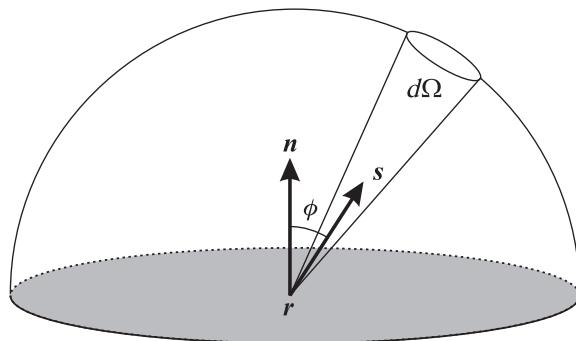


Fig. 3.4

Illustrating the calculation of the spectral irradiance by integrating the spectral radiance over the hemisphere above the shaded horizontal surface. The scalar product $\mathbf{n} \cdot \mathbf{s}$ in [equation \(3.5\)](#) arises from the projection of the unit area perpendicular to \mathbf{s} in the direction \mathbf{n} of the normal.

$$F(\mathbf{r}, \mathbf{n}) = \int_0^\infty F_\nu(\mathbf{r}, \mathbf{n}) d\nu = \int_{2\pi} L(\mathbf{r}, \mathbf{s}) \mathbf{n} \cdot \mathbf{s} d\Omega(\mathbf{s}).$$

Its units are W m^{-2} .

It must be borne in mind that the irradiance has a specific direction associated with it; for example, if the surface in question (assumed for the present argument to be an imaginary, rather than a material, surface) is horizontal and the normal \mathbf{n} points upwards, then the irradiance under consideration (denoted by F^\uparrow) is associated with upward-moving photons. Conversely, the irradiance $F^\downarrow = F(\mathbf{r}, -\mathbf{n})$ is associated with downward-moving photons. If we require the net upward power per unit area, F_z say, then we must take the difference:

$$F_z = F^\uparrow - F^\downarrow.$$

In terms of electromagnetic quantities, this net upward irradiance equals the vertical component of the Poynting vector.

There is a simple relationship between the radiance and irradiance from an isothermal plane surface, at temperature T , that emits black-body radiation. Since the black-body radiation is isotropic, $L_\nu = B_\nu(T)$ is independent of \mathbf{s} and \mathbf{r} , so the hemispheric integral is straightforward. [Equation \(3.5\)](#) for the spectral irradiance becomes

$$F_\nu(\mathbf{r}, \mathbf{n}) = \int_{2\pi} L_\nu \mathbf{n} \cdot \mathbf{s} d\Omega(\mathbf{s}) = 2\pi B_\nu \int_0^{\pi/2} \cos \phi \sin \phi d\phi = \pi B_\nu(T), \quad (3.6)$$

where ϕ is the angle between \mathbf{s} and the normal \mathbf{n} (see [Figure 3.4](#)) so that $d\Omega = 2\pi \sin \phi d\phi$, since there is axisymmetry around the normal. Integrating over all ν we obtain the Stefan–Boltzmann law for the irradiance

$$F(\mathbf{r}, \mathbf{n}) = \pi \int_0^\infty B_\nu(T) d\nu = \sigma T^4, \quad (3.7)$$

using the frequency-integral analogue of [equation \(3.3\)](#).

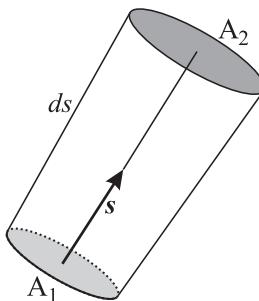


Fig. 3.5

A beam, or ‘pencil’, of radiation travelling a distance ds from surface A_1 of unit area to a surface A_2 . (Scattering of photons out of or into the beam is not indicated.) The area of A_2 is slightly greater than unity, owing to the divergence of photons into a solid angle $\Delta\Omega$ from each point of A_1 ; see Figure 3.3. However, to leading order, the volume of the pencil is still given by (unit area) \times ds .

3.2.2 Extinction and emission

Consider a beam of radiation of unit cross-sectional area, moving in a small range of solid angles $\Delta\Omega$ about the direction s ; see Figure 3.5. If the photons experience absorption or scattering in a small distance ds along the beam, due to the presence of a radiatively active gas,⁴ then the spectral radiance L_ν will be reduced. The physics of the process is complex; however, it may be summed up by **Lambert’s law**, which states that the fractional decrease of the spectral radiance is proportional to the mass of absorbing or scattering material encountered by the beam in a distance ds . Since the beam has unit cross-sectional area, this mass is $\rho_a ds$, where ρ_a is the density of the radiatively active gas, so

$$dL_\nu = -k_\nu(s)\rho_a(s)L_\nu(s)ds. \quad (3.8)$$

(The dependence of L_ν on the direction vector s is omitted here for clarity.) The quantity k_ν is called the **extinction coefficient**; it is the sum of an absorption coefficient a_ν and a scattering coefficient s_ν , defined in an obvious manner in terms of the contributions to dL_ν from absorption and scattering, respectively:

$$k_\nu = a_\nu + s_\nu. \quad (3.9)$$

The extinction coefficient k_ν generally depends on temperature and pressure, and can be regarded as calculable from detailed quantum mechanics or as an empirical quantity, to be derived from measurements. (Note that ρ_a is not generally the same as the total gas density ρ .)

If the gas is also emitting photons of frequency ν , an extra term must be added to the right-hand side of equation (3.8), to represent the additional power per unit area introduced into the beam. This term will also be proportional to the mass $\rho_a ds$, so it is convenient to write it as $k_\nu \rho_a J_\nu ds$, where $J_\nu(s)$ is called the **source function**. Including both extinction and

⁴ We refer to gases here, but the same considerations also apply to a gas containing a suspension of solid particles or liquid droplets.

emission we therefore obtain the **radiative-transfer equation** (also called **Schwarzschild's equation**)

$$\frac{dL_\nu}{ds} = -k_\nu \rho_a (L_\nu - J_\nu). \quad (3.10)$$

If k_ν , ρ_a and J_ν are given as functions of distance s , a formal solution of the radiative-transfer equation can be obtained as follows. First introduce the **optical path** χ_ν , defined by

$$\chi_\nu(s) = \int_{s_0}^s k_\nu(s') \rho_a(s') ds', \quad (3.11)$$

where s_0 is the start of the path; then [equation \(3.10\)](#) can be written as

$$\frac{dL_\nu}{d\chi_\nu} + L_\nu = J_\nu. \quad (3.12)$$

After multiplying by the integrating factor $\exp(\chi_\nu)$ we can integrate [equation \(3.12\)](#) to get

$$L_\nu e^{\chi_\nu} = \int J_\nu e^{\chi_\nu'} d\chi_\nu' + \text{constant}.$$

If the spectral radiance equals $L_{\nu 0}$ at the point s_0 then

$$L_\nu(s) = \int_0^{\chi_\nu} J_\nu(\chi') e^{-(\chi_\nu - \chi')} d\chi' + L_{\nu 0} e^{-\chi_\nu}. \quad (3.13)$$

Note that, in the absence of emission ($J_\nu = 0$), the spectral radiance falls exponentially, decreasing by a factor of e over a distance corresponding to unit optical path. (This exponential decay is known as **Beer's law**.) A region is said to be **optically thick** at a frequency ν if the total optical path χ_ν through the region is much greater than 1, and **optically thin** if the total optical path is much less than 1. A photon is likely to be absorbed or scattered within an optically thick region, but is likely to traverse an optically thin region without absorption or scattering.

As a simple example, suppose that the extinction coefficient k_ν and the density ρ_a of the radiatively active gas are both constant and take $s_0 = 0$. Then from [equation \(3.11\)](#) the optical path is proportional to the distance s , $\chi_\nu = k_\nu \rho_a s$, so

$$L_\nu(s) = k_\nu \rho_a \int_0^s J_\nu(s') e^{-k_\nu \rho_a (s-s')} ds' + L_{\nu 0} e^{-k_\nu \rho_a s}. \quad (3.14)$$

The radiance $L_\nu(s)$ reaching s thus has a simple interpretation, as follows. The second term on the right-hand side of [equation \(3.14\)](#) represents the radiance at the starting point $s = 0$, attenuated by an exponential factor due to extinction over the distance s , while the integral represents the sum of contributions emitted from elements ds' at different distances s' along the path, each attenuated by the factor $\exp[-k_\nu \rho_a (s - s')]$ due to extinction over the remaining distance $s - s'$. (See [Figure 3.6](#).)

Under local thermodynamic equilibrium conditions (see [Section 3.1.2](#)), in the absence of scattering, the source function equals the black-body spectral radiance, i.e. the Planck function, [\(3.1\)](#). This can be shown by using Kirchhoff's law (see [Section 3.1.1](#)), which holds under LTE conditions, as follows. The radiance emitted from a mass $\rho_a ds$ of gas in the beam is $k_\nu \rho_a ds J_\nu$ and the radiance absorbed is $k_\nu \rho_a ds L_\nu$; cf. [equation \(3.10\)](#). Hence

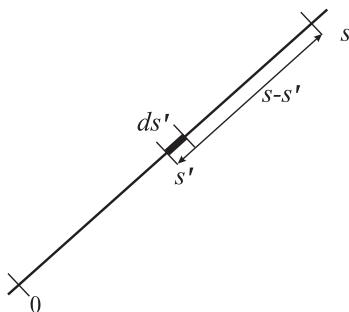


Fig. 3.6

Illustrating the terms in [equation \(3.14\)](#).

the spectral emittance, the ratio of the emitted radiance to the radiance emitted by a black body, is $\varepsilon_\nu = k_\nu \rho_a ds J_\nu / B_\nu$. The spectral absorptance, the fraction of incident radiance that is absorbed, is $\alpha_\nu = k_\nu \rho_a ds L_\nu / L_\nu = k_\nu \rho_a ds$, neglecting scattering. However, Kirchhoff's law states that $\varepsilon_\nu = \alpha_\nu$ and hence $J_\nu = B_\nu$.

3.2.3 The diffuse approximation

In radiative calculations we can often assume that the properties of the atmosphere and the radiation depend only on the vertical coordinate z .⁵ This is the **plane-parallel atmosphere** assumption and we shall make it from now on. The net irradiance (taking account of photons moving in opposite directions) is then vertical and equal to $F_z(z)$, as defined in [Section 3.2.1](#). Calculation of the upward and downward irradiances involves the integration, over the solid angle, of radiances such as that in [equation \(3.13\)](#), with the direction vector s re-inserted. This integration, though conceptually straightforward, is quite complicated in practice. However, a simple and surprisingly accurate alternative is to use the **diffuse approximation**. This states that we can replace the radiative-transfer [equation \(3.12\)](#), for calculating the spectral radiances L_ν along a set of slanting downward paths (with $J_\nu = B_\nu$, assuming LTE conditions), by the single equation

$$\frac{dF_\nu^\downarrow}{d\chi_\nu^*} + F_\nu^\downarrow = \pi B_\nu \quad (3.15)$$

for the downward spectral irradiance F_ν^\downarrow along a *vertical* path, where

$$\chi_\nu^* \approx 1.66 \chi_\nu. \quad (3.16)$$

The quantity χ_ν is the optical path measured downwards from the top of the atmosphere and is called the **optical depth**; see [equation \(3.29\)](#). The quantity χ_ν^* is a scaled optical

⁵ An important exception is when horizontal inhomogeneities due to clouds are present.

depth. The factor π on the right-hand side of equation (3.15) arises from the calculation of the black-body irradiance, as in equation (3.6). A similar equation, but with a change in sign of χ_v^* , holds for the upward spectral irradiance; see equation (3.44a).

3.3 Basic spectroscopy of molecules

3.3.1 Vibrational and rotational states

It was mentioned in Section 3.1 that the absorption and emission of thermal photons depend on energy differences between vibrationally or rotationally excited states of atmospheric molecules. The detailed spectroscopy of atmospheric molecules is a highly complex subject, well beyond the scope of this book. However, the basic principles are comparatively simple and will be sketched out in this section. As an illustration, we consider the vibrational and rotational states of a diatomic molecule, composed of an atom of mass m_1 and an atom of mass m_2 ; see Figure 3.7.

The classical picture of the vibration of this molecule assumes that the atoms are bound by a light ‘spring’, of spring constant K , that resists deviations of the distance x between the atoms from its equilibrium value, x_0 say. It is a straightforward exercise in classical mechanics to show that the frequency of oscillation of the system is ν_0 , given by

$$2\pi\nu_0 = \left(\frac{K}{m_r}\right)^{1/2}, \quad \text{where} \quad m_r = \frac{m_1m_2}{m_1 + m_2}.$$

The quantity m_r is called the **reduced mass**. The quantum-mechanical theory of the harmonic oscillator requires that we insert the potential function

$$V(x) = \frac{1}{2}K(x - x_0)^2$$

for this system into Schrödinger’s equation; there results an infinite set of energy levels, given by

$$E_v = \hbar\nu_0\left(v + \frac{1}{2}\right), \quad \text{where} \quad v = 0, 1, 2, \dots \quad (3.17)$$

Here v is the **vibrational quantum number** and takes integer values. The levels are **non-degenerate**; that is, there is only one state corresponding to each energy value E_v .

The absorption and emission of photons occurs when there is a time-dependent interaction between a molecule and an electromagnetic field, described quantum-mechanically by an interaction Hamiltonian. For the strongest vibrational interactions this Hamiltonian takes the form $-\mathbf{p} \cdot \mathbf{E}(t)$, where \mathbf{p} is the electric dipole moment of the molecule and \mathbf{E} is the

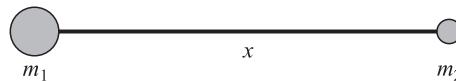


Fig. 3.7

A simple representation of a diatomic molecule, composed of atoms of masses m_1 and m_2 , separated by a distance x .

electric field in the neighbourhood of the molecule. For such a ‘dipole transition’ to occur, the dipole moment must change between the initial and final state. The quantum-mechanical **selection rule** for a diatomic molecule states that, in a dipole transition from one vibrational state to another, v can only increase or decrease by unity ($\Delta v = \pm 1$), corresponding to an energy change $\Delta E = \pm h\nu_0$.

We now consider the rotation of the same diatomic molecule, ignoring vibrations so that $x = x_0$. Classically, the molecule has a moment of inertia $I = m_r x_0$ and can rotate with any angular momentum. The quantum-mechanical analysis of the system is similar to that for the angular structure of the hydrogen atom, for example. It is found that the only allowed values of the squared angular momentum are $\hbar^2 J(J+1)$ where J is an integer, the **rotational quantum number**, and $\hbar = h/(2\pi)$. The corresponding energy levels are given by

$$E_J = \frac{1}{2I} J(J+1) \hbar^2, \quad \text{where } J = 0, 1, 2, \dots; \quad (3.18)$$

these levels are **degenerate**, with $2J + 1$ states corresponding to each energy value E_J . Rotational dipole transitions can occur only if the molecule has a nonzero permanent electric dipole moment. The selection rule in this case is $\Delta J = \pm 1$.

Equations (3.17) and (3.18) illustrate the fact that, when quantum mechanics is taken into account, only discrete (quantised) energy levels are allowed for vibrational and rotational states of the simple diatomic molecule. Furthermore, transitions between these quantised levels, associated with absorption or emission of photons, must lead to energy differences ΔE and hence photon frequencies $\nu = \Delta E/h$ that are also discrete in nature.

The treatment of molecules with more than two atoms is more complex; however, for future reference we mention that the vibrational normal modes of linear, symmetric triatomic molecules (such as carbon dioxide) and certain triatomic molecules that are not linear (such as water vapour and ozone) take fairly simple forms, as illustrated in Figure 3.8. These normal modes are designated by the labels ν_1 , ν_2 and ν_3 , as shown, and corresponding vibrational quantum numbers v_1 , v_2 and v_3 are used to represent the different excitations of each mode.

Wavelengths $hc/\Delta E_v$ associated with transitions between pure vibrational states are typically between about 1 and 20 μm and hence mostly in the thermal infra-red. Wavelengths $hc/\Delta E_J$ associated with transitions between pure rotational states would be of order 10^2 – 10^4 μm , in the far infra-red and microwave regions. In general, however, both vibrational and rotational energies change during a transition. The fact that the rotational energy differences are so much smaller than the vibrational energy differences means that, in terms of atmospheric spectra, rotational effects lead to ‘fine-structure splitting’ of vibrational spectral lines. Transitions from rotational sub-levels of one vibrational level to rotational sub-levels of another vibrational level therefore give rise to complex **vibration–rotation** spectral bands in the infra-red.

A very important consequence of the results stated above is that dipole transitions cannot occur for diatomic molecules with identical nuclei, such as the two most prevalent atmospheric gases N_2 and O_2 . Symmetric molecules of this type have zero dipole moments and hence cannot interact directly with thermal infra-red radiation, through either vibrational or rotational transitions. The same is true of transitions involving only excitations of the

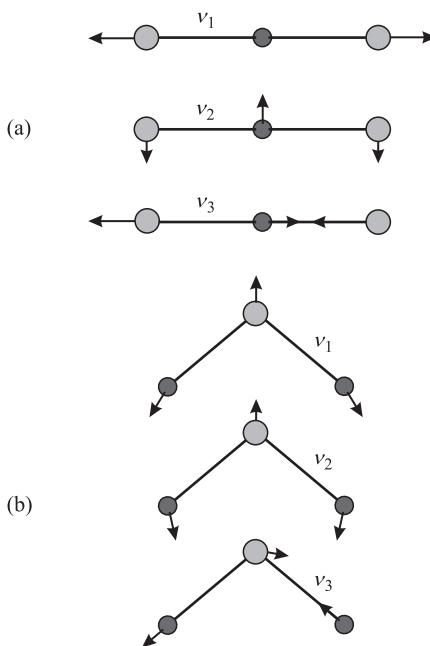


Fig. 3.8

Schematic illustration of vibrational normal modes of triatomic molecules. (a) A linear triatomic molecule such as CO_2 , showing the symmetric ν_1 and ν_2 modes and the asymmetric ν_3 mode. (Note that two independent forms of the ν_2 mode occur, one with displacements in the plane of the paper, as shown, and one with displacements perpendicular to the plane of the paper.) (b) A non-linear triatomic molecule such as H_2O . Adapted from Herzberg (1945).

‘symmetric stretch’ ν_1 mode of CO_2 (see Figure 3.8(a)), which also has a zero dipole moment. However, transitions involving triatomic modes such as the others illustrated in Figure 3.8 do lead to changes in electric dipole moment and can therefore interact strongly with thermal infra-red radiation, subject to appropriate selection rules. This explains why the major species N_2 and O_2 are effectively transparent to thermal infrared radiation, and why minor species⁶ such as CO_2 , H_2O and O_3 act as ‘greenhouse gases’ and play a dominant role in the Earth’s radiation balance and hence in determining the climate: see Chapter 8.

Wavelengths associated with transitions between electronically excited states are generally less than $1 \mu\text{m}$ and lie in the ultra-violet and visible region. Thus, in the case of solar photons, we must also consider the electronically excited states of molecules. This again is a complex subject, but it can be described qualitatively in the case of a diatomic molecule, by using an energy diagram, such as Figure 3.9. The lower curved line shows the potential energy of the ground state versus the internuclear distance. It takes the form of a potential well resulting from electric repulsion at short distances and attraction at larger distances. The potential energy becomes constant at large enough distances, corresponding to dissociation of the molecule AB into two ground-state atoms, A and B. Within the

⁶ Minor species are gases that are present in much smaller quantities than molecular nitrogen (N_2) and molecular oxygen (O_2); see Section 2.2 and Table 2.1.

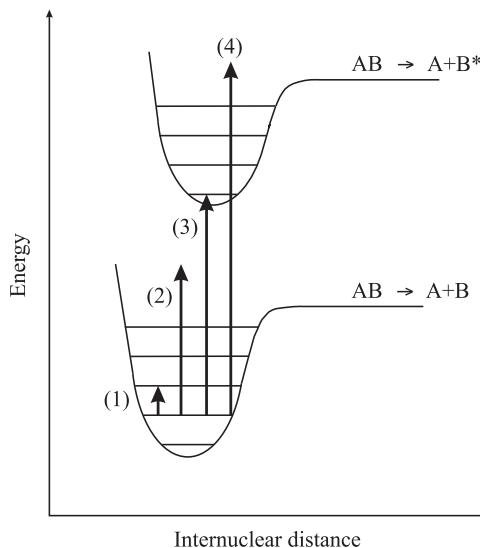


Fig. 3.9

Schematic energy diagram for a diatomic molecule. See text for details.

potential well there are a number of vibrational energy levels, denoted by the equispaced horizontal lines. (Near the bottom of the well, the potential energy varies approximately quadratically with distance, so that the simple theory leading to equation (3.17) applies there.) A similar potential well applies to the first electronically excited state; however, in this case dissociation leads to one ground-state atom (A say) and one electronically excited atom (B^* say).

Several transitions are shown in Figure 3.9 by vertical arrows. These are (1) the absorption of a photon, giving rise to an infra-red transition from one vibrational level to another vibrational level in the ground-state molecule; (2) photo-dissociation to two ground-state atoms; (3) absorption giving rise to a transition from a vibrational level of the ground state to a vibrational level in the excited state; and (4) photo-dissociation to one ground-state atom and one excited atom. Note that transitions (1) and (3) lead to discrete energy changes ΔE whereas (2) and (4) lead to continuous energy changes.

3.3.2 Line shapes

In this section we again neglect scattering, so that only absorption contributes to the extinction coefficient k_ν , defined in Section 3.2.2. The implication of the simple pictures of vibrational and rotational transitions given in the previous section is that k_ν can generally be expected to vanish, except at certain discrete frequencies ν_n associated with transitions, for which k_ν is large; these frequencies correspond to **spectral lines**. This suggests a model

$$k_\nu = \sum_n S_n \delta(\nu - \nu_n), \quad (3.19)$$

where the S_n are constants, called the **line strengths**, and $\delta(\cdot)$ is the Dirac delta function.

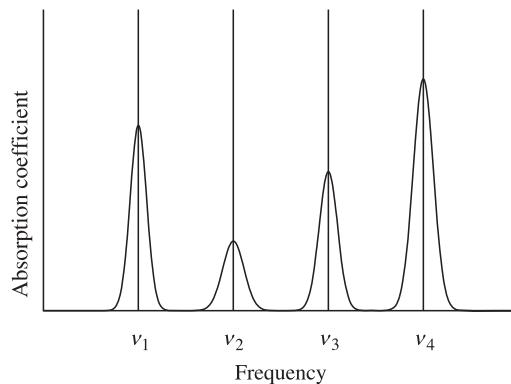


Fig. 3.10

A schematic diagram of some broadened spectral lines centred at four frequencies, v_1 , v_2 , v_3 and v_4 . Unbroadened lines would consist of sharp peaks at these frequencies, as indicated by the vertical lines.

However, this model of sharp, delta-function, peaks in the extinction coefficient is an idealisation. In practice several physical effects lead to a broadening of the sharp peaks (see Figure 3.10), so that equation (3.19) is replaced by

$$k_v = \sum_n S_n f_n(v - v_n),$$

where the f_n are **line-shape functions**, each normalised such that

$$\int_{-\infty}^{\infty} f_n(v - v_n) dv = 1. \quad (3.20)$$

There are two main cases.

(a) Collisional or natural broadening

The energy levels have finite lifetimes for one of two reasons:

- there is a collisional lifetime τ_c (see Section 3.1.2);
- excited states have a natural lifetime τ_n with respect to spontaneous decay.

The finite lifetimes imply that the energy levels are not precisely defined, but rather have a spread of values of order \hbar/τ_c or \hbar/τ_n , respectively, according to Heisenberg's Uncertainty Principle. There is thus a spread of values of each ΔE and hence of each transition frequency v_n .

In the case of collisional broadening we obtain the **Lorentz line shape**, of the form

$$f(v - v_n) = \left(\frac{\gamma_L}{\pi} \right) \frac{1}{(v - v_n)^2 + \gamma_L^2}, \quad (3.21)$$

where $\gamma_L = (2\pi\tau_c)^{-1}$ is the half-width at half maximum of the line. Note that f satisfies equation (3.20). The derivation of equation (3.21) is quite complicated; the proof is omitted from this book, but is given in advanced texts on quantum mechanics or spectroscopy.

We can get τ_c and hence γ_L from kinetic theory, in terms of the mean molecular velocity (which is proportional to $T^{1/2}$), number density (proportional to p/T) and collision cross-section, giving the following temperature and pressure dependence of γ_L :

$$\gamma_L \propto p T^{-1/2}. \quad (3.22)$$

A similar calculation applies to natural broadening, but in this case $\gamma_L = (2\pi\tau_n)^{-1}$. In the infra-red, the corresponding line width is much less than that due to collisional broadening and than that due to the Doppler broadening discussed below, so natural broadening can usually be neglected. However, natural broadening can be important in the ultra-violet.

(b) Doppler broadening

The Doppler shift of a spectral line at a frequency $\nu = \nu_0$ due to a molecular velocity u of the emitter away from the observer is $\nu - \nu_0 = (u/c)\nu_0$, assuming that $u \ll c$. However, the velocities follow the Maxwell–Boltzmann distribution; the probability that the speed of a molecule lies between u and $u + du$ is $P(u) du$, where

$$P(u) = \left(\frac{m}{2\pi kT} \right)^{1/2} \exp \left(-\frac{mu^2}{2k_B T} \right)$$

and m is the molecular mass. The spectral line shape, rather than being a delta function, becomes the convolution of $P(u)$ with a delta function:

$$\int_{-\infty}^{\infty} P(u) \delta\left(\nu - \nu_0 - \frac{u}{c}\nu_0\right) du \propto \exp \left(-\frac{mc^2(\nu - \nu_0)^2}{2k_B T \nu_0^2} \right).$$

We then get the **Doppler line shape**

$$k_\nu = \frac{S}{\gamma_D \sqrt{\pi}} \exp \left(-\frac{(\nu - \nu_0)^2}{\gamma_D^2} \right),$$

on normalising such that $S = \int_{-\infty}^{\infty} k_\nu d\nu$, where

$$\gamma_D = \frac{\nu_0}{c} \left(\frac{2k_B T}{m} \right)^{1/2}.$$

The half-width at half maximum is $\gamma_D (\ln 2)^{1/2}$, and is proportional to $T^{1/2}$ but independent of p .

The Lorentz and Doppler line shapes are shown in Figure 3.11; also shown is the Voigt line shape, which applies when both collisional and Doppler broadening are important. The fact that γ_L is proportional to pressure (see equation (3.22)) means that collisional broadening dominates at low altitudes (<30 km), in the infra-red. Higher up, Doppler broadening becomes significant in the visible and ultra-violet.

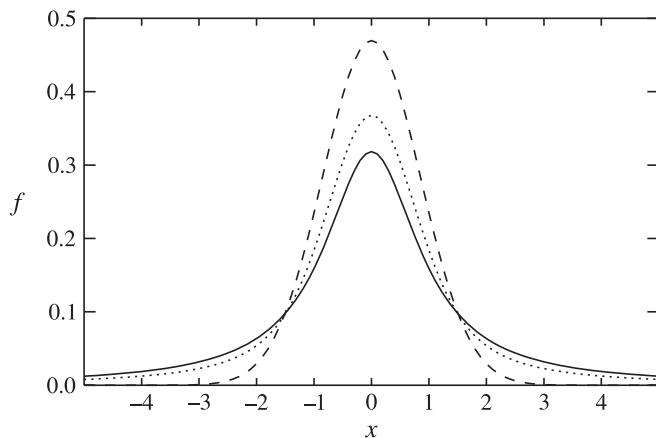


Fig. 3.11 Illustrating the Lorentz (solid), Doppler (dashed) and Voigt (dotted) line shapes as a function of $x = (v - v_0)/\alpha$, where α is the half-width at half maximum appropriate for each shape. The curves are normalised such that the area under each is the same.

3.4 Transmittance

An important quantity that arises in the solution of the radiative-transfer equation in Section 3.2.2 and the calculation of heating rates in Section 3.6 is the fraction of the spectral radiance leaving one point that arrives at another point. This fraction is represented by the **transmittance** or **transmission function**. For a parallel beam leaving point s_1 and arriving at point s_2 (or vice versa), the **spectral transmittance** is

$$\mathcal{T}_v(s_1, s_2) = \exp\left(-\left|\int_{s_1}^{s_2} k_v(s) \rho_a(s) ds\right|\right) = \exp[-|\chi_v(s_2) - \chi_v(s_1)|]; \quad (3.23)$$

cf. equations (3.11) and (3.13). The modulus signs express the fact that the fraction is independent of whether the radiation goes from s_1 to s_2 or from s_2 to s_1 and ensure that the fraction is ≤ 1 . If scattering is neglected, the extinction coefficient k_v depends on s through its temperature and pressure dependences. However, if the variation of k_v along the path can be ignored, such as, for example, for measurements under laboratory conditions at fixed p and T , then equation (3.23) gives

$$\mathcal{T}_v(s_1, s_2) = \exp[-k_v(p, T) u_a(s_1, s_2)], \quad (3.24)$$

where

$$u_a(s_1, s_2) = \left| \int_{s_1}^{s_2} \rho_a(s) ds \right|.$$

Here u_a is the mass of absorber gas, per unit transverse cross-sectional area, in the path. If the absorber density is also constant along the path, then $u_a = \rho_a l$, where l is the length of the path.

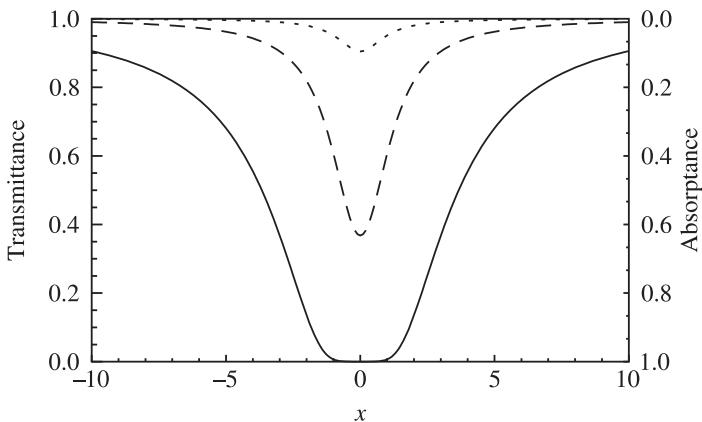


Fig. 3.12 The transmittance T_ν and absorptance A_ν as functions of $x = (\nu - \nu_0)/\gamma_L$ for a Lorentz line, for three values of the quantity $q = u_a S/(\gamma_L \pi)$: $q = 10$, solid curve; $q = 1$, dashed curve; and $q = 0.1$, dotted curve.

Figure 3.12 shows examples of the transmittance T_ν as a function of ν for three values of $q = u_a S/(\gamma_L \pi)$ for a Lorentz line at fixed p and T . Also indicated is the **absorptance** $A_\nu = 1 - T_\nu$. Note that, for large amounts of absorber ($q \gg 1$, solid curve) the transmittance is effectively zero near the line centre, whereas for small amounts of absorber ($q \ll 1$, dotted curve) the transmittance is close to 1 for all frequencies.

It is often useful to average the transmittance over a spectral band, of width $\Delta\nu_r$, say, perhaps containing many spectral lines, to get the **band transmittance**

$$\bar{T}_r = \frac{1}{\Delta\nu_r} \int_{\Delta\nu_r} T_\nu d\nu. \quad (3.25)$$

Two other useful quantities are the **band absorptance**

$$\bar{A}_r = 1 - \bar{T}_r,$$

and the **equivalent width or integrated absorptance**

$$\mathcal{W}_r = \int_{\Delta\nu_r} (1 - T_\nu) d\nu = \Delta\nu_r (1 - \bar{T}_r) = \Delta\nu_r \bar{A}_r. \quad (3.26)$$

If the integration is over a single broadened spectral line, then \mathcal{W}_r represents the width of a rectangular line shape within which total absorption takes place, which has the same area as the actual line.

Note that, for small amounts of absorber (again at fixed p and T along the path), we have $T_\nu = e^{-k_\nu u_a} \approx 1 - k_\nu u_a$, so that

$$\mathcal{W}_r \approx \int_{\Delta\nu_r} k_\nu u_a d\nu = S u_a, \quad (3.27)$$

using $S = \int_{\Delta\nu_r} k_\nu d\nu$. This is called the **weak-line approximation**. At the opposite extreme, for large amounts of absorber, the equivalent width can be calculated explicitly for a single Lorentz or Doppler line shape, assuming that k_ν is constant, to obtain the **strong-line**

approximations $\mathcal{W} \approx 2(Su_a\gamma_L)^{1/2}$ for a Lorentz line and $\mathcal{W} \approx 2\gamma_D\{\ln[Su_a/(\gamma_D\sqrt{\pi})]\}^{1/2}$ for a Doppler line; see Problems 3.5 and 3.7.

3.5 Absorption by atmospheric gases

3.5.1 The solar spectrum

Figure 3.13 shows the spectral irradiance of solar radiation at the top of the atmosphere (TOA) and at sea level, compared with the black-body spectral irradiance at a typical solar

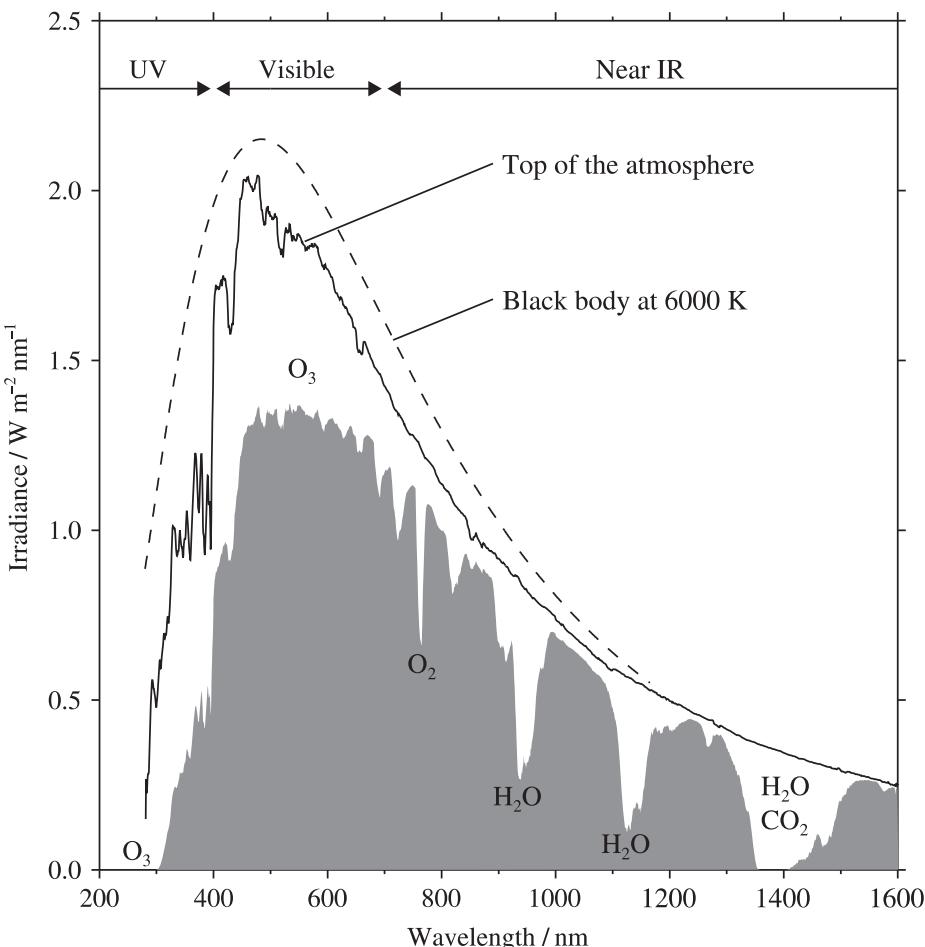


Fig. 3.13 The irradiance spectrum of solar radiation at the top of the atmosphere (solid line) and at sea level (shaded), compared with the black-body irradiance spectrum (dashed line), given by $B_\lambda(T)$ times the solid angle subtended by the Sun, for $T = 6000 \text{ K}$. Gases responsible for the most prominent absorption features are indicated. The approximate wavelength regions of ultra-violet, visible and near infra-red radiation are also shown. Plotted using data from the ASTM G173-03 Reference Spectra, see <http://rredc.nrel.gov/solar/spectra/am1.5/>.

photospheric temperature of 6000 K. It is seen that the TOA irradiance is fairly close to this black-body irradiance at most wavelengths. However, absorption and scattering in the atmosphere cause significant differences between the TOA irradiance and the sea-level irradiance, with especially large deviations (apparent in the sharp dips in the sea-level curve) at certain wavelengths, which are identified with particular absorbing gases, notably ozone (O_3) in the ultra-violet and visible, and carbon dioxide (CO_2) and water vapour (H_2O) in the infra-red. (The ‘red bands’ of O_2 , on the boundary between the visible and the infra-red, are associated with electronic transitions.) We investigate atmospheric absorption in more detail in the next two sections.

3.5.2 Infra-red absorption

The absorption of infra-red radiation by the six most significant gaseous absorbers is conveniently summarised in [Figure 3.14](#). This figure shows the transmittance for a vertical beam passing through the whole atmosphere, as a function of wavelength. The gases shown are all minor constituents, and all but ozone (O_3) are concentrated mainly in the troposphere. (As mentioned in [Section 3.3.1](#), the major constituents, N_2 and O_2 , are not strongly radiatively active in the infra-red.) Since, on the scale of the diagram, much of the fine structure associated with individual spectral lines is not shown, the diagram can be regarded as plots of the band transmittance \bar{T}_r (equal to $1 - \bar{A}_r$, where \bar{A}_r is the band absorptance) between the top of the atmosphere and the ground, corresponding to band widths $\Delta\nu_r$ associated with wavelength differences $\sim 0.1\text{ }\mu\text{m}$.

The bottom panel of [Figure 3.14](#) shows the total long-wave absorptance due to all gases. There is a broad region from about 8 to 13 μm , called the **atmospheric window**, within which absorption is weak, except for a band near 9.6 μm associated with O_3 .

Water vapour (H_2O) absorbs strongly over a wide band of wavelengths near 6.3 μm (associated with transitions involving the ν_2 vibrational mode: see [Figure 3.8](#)) and over a narrower band near 2.7 μm (associated with the ν_1 and ν_3 vibrational modes). At longer wavelengths, especially beyond 16 μm , rotational transitions of H_2O become important, leading to strong absorption.

Carbon dioxide (CO_2) is a strong absorber in a broad band near 15 μm , associated with the vibrational ν_2 ‘bending’ mode, and in a narrower band near 4.3 μm , associated with the ν_3 ‘asymmetric stretching’ mode. (The band near 2.7 μm has a more complex origin.)

Ozone (O_3) absorbs strongly near 9.6 μm (associated with the ν_1 and ν_3 vibrational modes), in the atmospheric window. Since the other gases do not absorb significantly in this spectral region, ozone (which is mainly concentrated in the stratosphere) can therefore exchange radiation with the lower atmosphere; see [Section 3.6.4](#).

[Figure 3.14](#) gives information on the absorption of infra-red radiation over the total depth of the atmosphere, but not directly about the way in which the absorption varies with altitude. Moreover, it must be remembered that atmospheric gases also emit infra-red radiation and that this emission also varies with altitude. The vertical profiles of the absorption and emission are required in the calculation of the resulting heating and cooling; this is discussed in [Section 3.6.3](#).

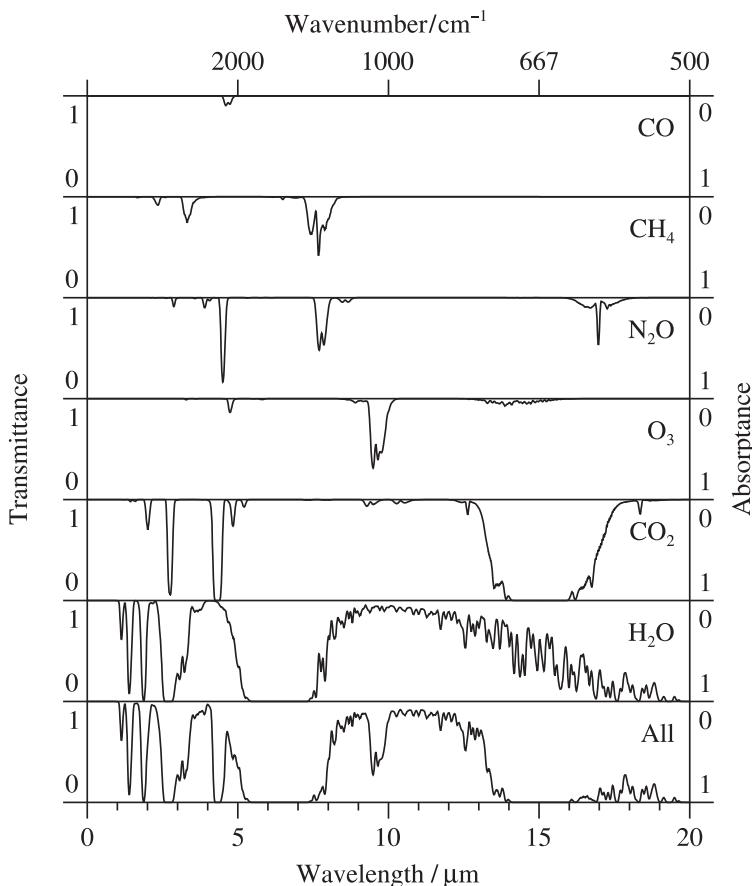


Fig. 3.14

Infra-red absorption spectra for six strongly absorbing gases and for the six gases combined, for a vertical beam passing through the atmosphere, in the absence of clouds. Drawn from data supplied by Dr A. Dudhia.

3.5.3 Ultra-violet absorption

In the ultra-violet, the main absorbers are molecular oxygen (O_2) and ozone. Absorption at these wavelengths is often depicted in terms of the **absorption cross-section** σ_ν , which is equal to the absorption coefficient a_ν times the molecular mass. Unlike in the infra-red, we must take account of electronic transitions, as well as vibrational and rotational transitions, when considering absorption at discrete wavelengths; moreover, photo-dissociation and photo-ionisation (see Section 3.1) lead to important **continuum absorption**, i.e. absorption over a continuous range of wavelengths, rather than at discrete wavelengths.

The absorption cross-section for O_2 (Figure 3.15) has large values due to ionisation at wavelengths below 100 nm; in the range 100–130 nm there are irregular bands of unknown origin. The Schumann–Runge continuum, in the range 130–175 nm, is due to the dissociation $O_2 \rightarrow O(^3P) + O(^1D)$, in which one oxygen atom remains in the ground

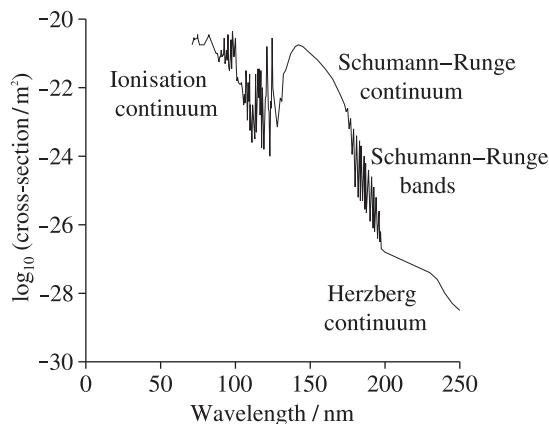


Fig. 3.15 General shape of the absorption cross-section as a function of wavelength for O_2 . Adapted from Figure 4.30 of Brasseur and Solomon (2005).

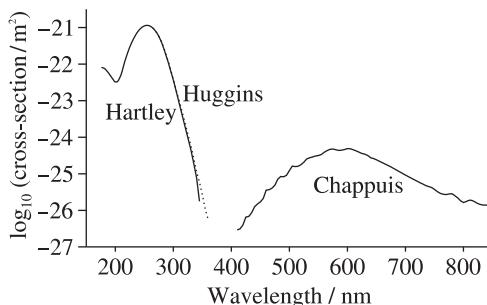


Fig. 3.16 The absorption cross-section as a function of wavelength for O_3 . Details of the fine structure of the Huggins band have been suppressed. In the Huggins band the solid line corresponds to a temperature of 203 K and the dashed line to a temperature of 273 K.

‘triplet-P’ state and the other goes to the excited ‘singlet-D’ state. The Schumann–Runge bands, in the range 175–200 nm, are associated with an electronic transition and superimposed vibrational transitions. The Herzberg continuum is found in the range 200–242 nm. At 242 nm dissociation into two ground-state oxygen atoms occurs; although this is an insignificant absorption feature, it is very important in the formation of ozone (see Section 6.5.1). A further electronic transition gives rise to the weak Herzberg bands in the range 242–260 nm.

The ozone absorption cross-section (Figure 3.16) exhibits two continua in the ultra-violet and also one in the visible and near infra-red, all due to photo-dissociation: the Hartley band in the range 200–310 nm, the Huggins bands in the range 310–350 nm and (in the visible and near infra-red) the Chappuis bands in the range 400–850 nm. Although the absorption cross-section for the Chappuis bands is much smaller than those for the Hartley and Huggins bands, the Chappuis bands are important since they occur near the peak of the solar spectrum and absorb in the troposphere and lower stratosphere. (The absorption features due to the Hartley, Huggins and Chappuis bands are indicated by the O_3 symbols in Figure 3.13.) Shorter-wavelength (more energetic) radiation is almost absent at these

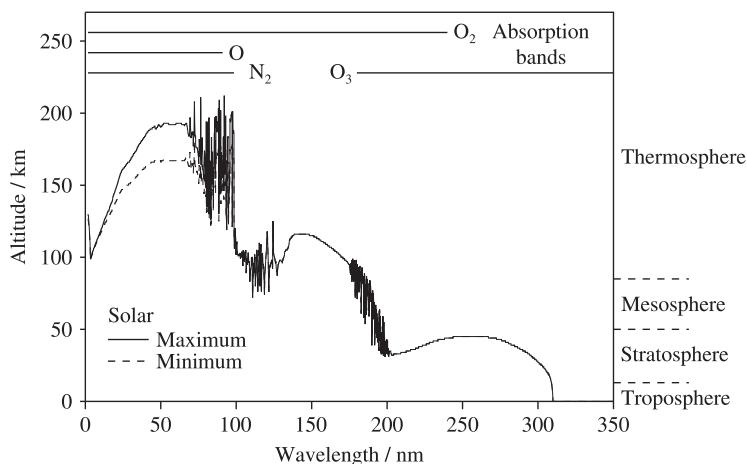


Fig. 3.17 The altitude of unit optical depth for vertical solar radiation. The principal absorption bands are shown. Adapted from Meier (1991); an early version of this figure appeared in Herzberg (1965). Figure courtesy of Dr J. Lean and Dr R. Meier.

levels, since it is mostly absorbed higher up. This is demonstrated by Figure 3.17, which shows the altitude of unit optical depth (the peak of the Chapman layer in absorption; see Figure 3.18) as a function of wavelength.

The heating of the atmosphere due to absorption of ultra-violet radiation is discussed in Section 3.6.2. In contrast to the infra-red, there is no significant emission from atmospheric gases in the ultra-violet, since the black-body spectral radiance at terrestrial temperatures is so small there; see Figures 3.1 and 3.2.

3.6 Heating rates

3.6.1 Basic ideas

One of the main goals of radiative calculations is to obtain radiative-heating rates throughout the atmosphere. For this one requires knowledge of the heating due to absorption of solar (short-wave) photons and the heating and cooling due to absorption and emission of thermal (long-wave) photons. In this section we consider some basic ideas; these are applied to solar and thermal radiation in later sections.

Consider a horizontal slab of atmosphere, of horizontal area A , at height z and of thickness Δz , and make the plane-parallel atmosphere assumption. The net upward power entering the bottom of the slab is $AF_z(z)$ and the net upward power emerging at the top is $AF_z(z + \Delta z)$. The loss of radiative power within the volume $A \Delta z$ of the slab is therefore $A [F_z(z) - F_z(z + \Delta z)] \approx -(A \Delta z) dF_z/dz$. This loss of radiative power implies that radiative **diabatic heating** (see Sections 2.4 and 4.10) of the slab is occurring at a rate of $-dF_z/dz$ per unit volume or

$$Q = -\frac{1}{\rho(z)} \frac{dF_z}{dz} \quad (3.28)$$

per unit mass, where ρ is the density of air. The units of Q are W kg^{-1} ; often the quantity Q/c_p arises in calculations of the dynamical effects of radiative heating, where c_p is the specific heat capacity at constant pressure (cf. [equation \(4.35\)](#)), and this quantity has units K s^{-1} . Since F^\uparrow and F^\downarrow both involve integration of spectral irradiances over frequency, F_z ($= F^\uparrow - F^\downarrow$) and Q also comprise contributions from different frequency bands.

3.6.2 Short-wave heating

Consider the diabatic heating rate per unit volume, ρQ_v^{sw} , produced by absorption of short-wave solar radiation of frequency v by a gas of density $\rho_a(z)$ and extinction coefficient $k_v(z)$; scattering will be neglected. Assuming that the Sun is directly overhead, the appropriate optical path for each frequency is the **optical depth**, measured vertically downwards from the top of the atmosphere (taken to be $z = \infty$),

$$\chi_v(z) = \int_z^\infty k_v(z') \rho_a(z') dz'. \quad (3.29)$$

(For the direct solar beam we do not integrate over solid angle, so the diffuse approximation of [Section 3.2.3](#) is not made.) By analogy with the second term of [equation \(3.13\)](#), the downward irradiance of the solar radiation is given by

$$F_v^\downarrow(z) = F_{v\infty}^\downarrow e^{-\chi_v(z)}, \quad (3.30)$$

where $F_{v\infty}^\downarrow$ is the downward solar irradiance at the top of the atmosphere and the exponential term $e^{-\chi_v(z)}$ equals the transmittance $T_v(z, \infty)$ between the top of the atmosphere and height z . Since scattering is neglected, the upward solar irradiance F_v^\uparrow must be zero and the net vertical irradiance at frequency v is

$$F_{zv}(z) = -F_{v\infty}^\downarrow e^{-\chi_v(z)}. \quad (3.31)$$

The contribution to the heating rate per unit volume from this frequency is therefore

$$\rho Q_v^{\text{sw}} = \frac{d}{dz} \left(F_{v\infty}^\downarrow e^{-\chi_v(z)} \right) = F_{v\infty}^\downarrow \left(-\frac{d\chi_v}{dz} \right) e^{-\chi_v(z)} = F_{v\infty}^\downarrow k_v(z) \rho_a(z) e^{-\chi_v(z)}.$$

Suppose now that the extinction coefficient k_v is independent of z and that the density of the absorber decays exponentially with height, $\rho_a(z) = \rho_a(0) e^{-z/H_a}$, where H_a is constant. Then the integral in [equation \(3.29\)](#) can be evaluated explicitly, giving

$$\chi_v(z) = H_a k_v \rho_a(0) e^{-z/H_a} = \chi_v(0) e^{-z/H_a};$$

this shows how the optical depth increases as the solar radiation penetrates downwards, i.e. as z decreases. Substitution into [equation \(3.31\)](#) then gives the vertical irradiance

$$F_{zv} = -F_{v\infty}^\downarrow \exp[-\chi_v(0) e^{-z/H_a}]$$

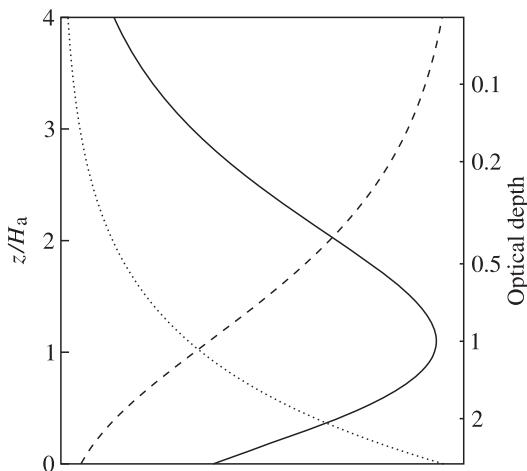


Fig. 3.18 Vertical profiles of the short-wave volume heating rate, $\rho Q_v^{\text{sw}}(z)$ (solid line), the negative of the vertical irradiance, $-F_{zv}(z)$ (dashed line) and the absorber density $\rho_a(z)$ (dotted line), for solar radiation at frequency ν in the simple example described in the text. The horizontal scales are arbitrary. The left-hand ordinate shows the height z , divided by H_a ; the right-hand ordinate shows the optical depth, $\chi_\nu(z)$. The optical depth of the ground $z = 0$ at frequency ν is arbitrarily chosen to be 3. The Sun is taken to be overhead. The absorber gas has a constant extinction coefficient and an exponentially decaying density $\rho_a \propto \exp(-z/H_a)$. Note that, in this example, the optical depth has the same exponential variation as the absorber density.

and differentiation gives the monochromatic volume heating rate,

$$\rho Q_v^{\text{sw}}(z) = F_{\nu\infty}^\downarrow k \rho_a(0) \exp\left(-\frac{z}{H_a} - \chi_\nu(0)e^{-z/H_a}\right). \quad (3.32)$$

Figure 3.18 shows the variation of the optical depth χ_ν , the vertical irradiance F_{zv} and the volume heating rate ρQ_v^{sw} , as functions of height in this simple example; note that the volume heating rate has a broad single peak. Differentiation of equation (3.32) with respect to z shows that ρQ_v^{sw} has a maximum at the height where $\chi_\nu(z) = 1$, that is, the height where the optical depth at frequency ν equals unity. (We assume that the absorber is sufficiently ‘optically thick’ for this level to occur above the ground.)

A vertical structure of the type given by equation (3.32) is said to exhibit a **Chapman layer**. The peaked shape of the Chapman layer in the heating rate can be interpreted as follows. At high levels, above the peak, there is a large vertical irradiance (since little absorption of the solar beam has yet occurred), but few absorber molecules; at low levels, below the peak, there is a small vertical irradiance (since much absorption has occurred) but many absorber molecules. In each case the heating rate is small. However, near the level of unit optical depth both the irradiance and the absorber density are significant and so the heating rate is comparatively large. Chapman layers also occur in other processes determined primarily by the absorption of radiation; an example is the photo-dissociation that contributes to the formation of the ‘ozone layer’; see Section 6.4.

3.6.3 Long-wave heating and cooling

We now consider the effects of thermal (long-wave) photons, allowing for both downward and upward paths. It can be shown, by solving the radiative transfer equation and integrating over solid angle, that the upward thermal irradiance at frequency ν and height z is

$$F_v^\uparrow(z) = \pi \int_0^z B_\nu(z') \frac{\partial T_\nu^*(z', z)}{\partial z'} dz' + \pi B_\nu(0) T_\nu^*(0, z). \quad (3.33)$$

Here $T_\nu^*(z', z)$ is the spectral transmittance, averaged over the upward hemisphere to take account of all slanting paths between heights z' and z , and $B_\nu(0)$ is the Planck function evaluated at the temperature of the Earth's surface. The assumption of LTE has been made, so that the source function $J_\nu = B_\nu$, and the Earth's surface has been assumed to radiate as a black body. Similarly, the downward irradiance is

$$F_v^\downarrow(z) = -\pi \int_z^\infty B_\nu(z') \frac{\partial T_\nu^*(z', z)}{\partial z'} dz';$$

there is no 'boundary' term here since the downward thermal irradiance at the top of the atmosphere is zero.

The net upward long-wave spectral irradiance is $F_{z\nu}(z) = F_v^\uparrow(z) - F_v^\downarrow(z)$ and from this the net long-wave diabatic heating rate per unit mass, Q_v^{lw} , can be calculated using equation (3.28). The resulting expression for Q_v^{lw} is quite complicated, but has a simple physical interpretation: namely, the net heating or cooling at a given level is due to the difference between the energy *gained* per unit time by absorption of photons from neighbouring levels and the Earth's surface, and the energy *lost* per unit time by emission of photons to neighbouring levels, the Earth's surface and space.

These heating and cooling terms can also be obtained directly. Consider for example the rate of loss of energy by a horizontal slab of atmosphere of thickness Δz and horizontal area A at height z by emission of photons to space – the **cooling-to-space** term. The derivation of the radiative-transfer equation (3.10) shows that the spectral power emitted in a vertical direction from this slab is $k_\nu \rho_a J_\nu A \Delta z$, where J_ν is the source function, equal to B_ν under LTE, as above. The fraction of this power that escapes to space is given by the transmittance $T_\nu(z, \infty) = \exp(-\int_z^\infty k_\nu \rho_a dz')$. Noting that

$$\frac{\partial T_\nu(z, \infty)}{\partial z} = k_\nu(z) \rho_a(z) T_\nu(z, \infty),$$

we find that the power escaping to space from the slab in a purely vertical beam is

$$B_\nu(z) \frac{\partial T_\nu(z, \infty)}{\partial z} A \Delta z.$$

Now, integrating over all slanting paths as above and replacing T_ν by T_ν^* , we obtain a contribution to the heating rate per unit mass

$$Q_v^{\text{cts}}(z) = -\frac{\pi B_\nu(z)}{\rho(z)} \frac{\partial T_\nu^*(z, \infty)}{\partial z}. \quad (3.34)$$

The factor π in this equation comes from integration over the upward hemisphere and the minus sign from the fact that the power loss to space implies a negative heating at height z . The other contributions to the heating rate at height z can be calculated in similar, but more complicated, ways; these must include the heating of the slab due to absorption of photons emitted at other levels. A useful simplification for some purposes is the **cooling-to-space approximation**, in which the loss of photon energy to space dominates the other contributions; therefore, under this approximation, $Q_v^{\text{lw}} \approx Q_v^{\text{cts}}$.

All gases that absorb and emit at frequency ν must in principle be included in T_ν^* . Then $Q_v^{\text{lw}}(z)$ must be integrated over all relevant frequencies to obtain the total long-wave cooling $Q^{\text{lw}}(z)$.

3.6.4 Net radiative heating rates

The short-wave and long-wave contributions to the diabatic heating rate Q can be computed using the principles described in the two previous sections, together with information on the atmospheric temperature structure and the concentration and spectroscopic characteristics of absorber gases. In this section we summarise the basic results of such computations.

Figure 3.19 shows the vertical profiles of the global-mean short-wave heating rate Q^{sw}/c_p and the long-wave cooling rate $-Q^{\text{lw}}/c_p$, in convenient units of K day^{-1} . The corresponding profiles of heating and cooling due to the most important atmospheric gases are also shown. It is clear that the total heating is approximately equal to the total cooling over much of the

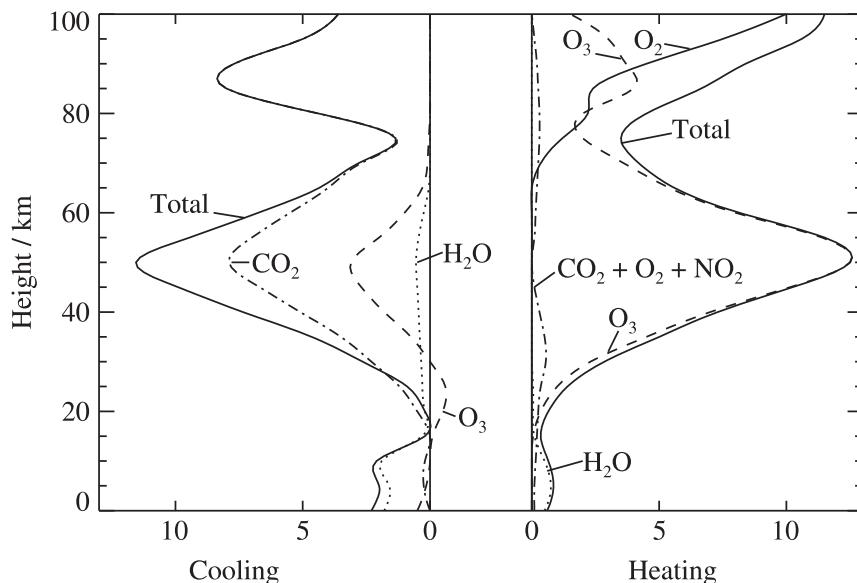


Fig. 3.19 Global-mean vertical profiles of the short-wave heating rate and the long-wave cooling rate, in K day^{-1} , including contributions from individual gases. Adapted from London (1980).

profile, except in the troposphere and above about 90 km. In the global mean, the middle atmosphere is therefore close to being in **radiative equilibrium**.

Below 80 km the short-wave heating rate is dominated by a Chapman-layer-like structure, centred near 50 km, due to absorption of solar radiation by ozone. The peak of the heating rate, at over 10 K day^{-1} , lies above the maxima both in the ozone density (near 22 km) and in the ozone mixing ratio (near 37 km). (The fact that this peak is above the maximum ozone density is consistent with [Figure 3.18](#) and is explained qualitatively by the argument at the end of [Section 3.6.2](#).) Below 15 km, in the troposphere, the main contribution to the heating rate is from water vapour, at about 1 K day^{-1} . Heating due to absorption by ozone and molecular oxygen is important between 80 and 100 km.

The peak in long-wave cooling near 50 km has significant contributions due to carbon dioxide and ozone, both cooling to space. The dominant wavelength bands involved are $15 \mu\text{m}$ for CO_2 and $9.6 \mu\text{m}$ for O_3 . A small amount of long-wave heating appears near 20 km, in the lower stratosphere, because of absorption by ozone of upwelling radiation from the troposphere at wavelengths near $9.6 \mu\text{m}$, in the atmospheric window. Tropospheric cooling is dominated by water vapour, at about 2 K day^{-1} .

[Figure 3.19](#) omits complications due to clouds and aerosols. Furthermore, the approximate radiative equilibrium in the global mean does not apply to individual latitudes and seasons. For example, in the winter stratosphere and the summer upper mesosphere there are big differences between Q^{sw} and $-Q^{\text{lw}}$. In such regions dynamical heat transport is also significant; see [Chapters 4 and 5](#).

It must be emphasised that the net radiative heating rate $Q = Q^{\text{sw}} + Q^{\text{lw}}$ should not be thought of as a pre-ordained heating, to which the atmospheric temperature and wind fields respond. One reason is that Q itself depends strongly on the temperature, especially through Q^{lw} . A highly simplified, but nevertheless illuminating, model is to regard Q at a point \mathbf{r} as a function of the local value of T and possibly of other variables. In radiative equilibrium $Q = 0$ by definition. Suppose that the corresponding radiative-equilibrium temperature field is $T_r(\mathbf{r})$; then $Q(T_r(\mathbf{r})) = 0$. If now the temperature deviates slightly from radiative equilibrium, $T = T_r + \delta T$, say, the net heating rate will also differ from zero:

$$Q(T_r + \delta T) \approx Q(T_r) + \delta T \frac{\partial Q}{\partial T} \Big|_{T=T_r} = \delta T \frac{\partial Q}{\partial T} \Big|_{T=T_r} = -c_p \frac{\delta T}{\tau_r} \quad (3.35)$$

say, since $Q(T_r) = 0$, where $\tau_r = -c_p(\partial Q/\partial T|_{T=T_r})^{-1}$. The net radiative heating rate is therefore proportional and opposite in sign to the deviation of the actual temperature from the radiative-equilibrium temperature. [Equation \(3.35\)](#) is one form of the **Newtonian cooling approximation** and the coefficient τ_r (which is positive in practice) is the **radiative relaxation time**. In this simple model (and in more realistic models) there is a kind of ‘radiative spring’, which tries to pull the temperature towards radiative equilibrium. This spring is opposed by other physical processes, especially dynamically induced heat transport that drives the atmosphere away from radiative equilibrium and thus forces the net radiative heating Q to be non-zero. In this sense we can regard the dynamics as driving the net radiative heating, rather than the other way round; see also [Section 9.5.2](#).

3.7 The greenhouse effect revisited

In [Section 1.3](#) we introduced some simple ideas concerning the **greenhouse effect**, by which radiation-absorbing gases in the atmosphere raise the surface temperature above the value that would occur if they were not present. We considered there a highly idealised atmosphere, consisting of a single homogeneous, isothermal layer in radiative equilibrium. In the present section we consider two models of a non-isothermal atmosphere, again assuming radiative equilibrium. The first model includes two homogeneous, isothermal atmospheric layers: an optically thin stratosphere on top of a troposphere. The second model uses some of the physics developed earlier in this chapter to study an inhomogeneous atmosphere in which the temperature varies continuously with height. It must be emphasised that these models are still very idealised; in particular they assume that heat transfer is by radiative processes only, thereby neglecting the heat transfer by fluid motions that was mentioned at the end of the previous section and that is an important feature of the real atmosphere. Such heat transfer can be brought about by processes such as convection and baroclinic waves (see [Section 5.7](#)) and may include latent heat effects.

3.7.1 Two-layer atmosphere in radiative equilibrium, including an optically thin stratosphere

This model is an extension of the single-layer model considered in [Section 1.3.1](#). It includes two atmospheric layers: the lower layer, at temperature T_{trop} , mimics the troposphere, and has transmittances \mathcal{T}_{lw} in the infra-red and \mathcal{T}_{sw} in the short-wave region, and emittance $1 - \mathcal{T}_{\text{lw}}$ in the infra-red, as in [Section 1.3.1](#). The upper layer, at temperature T_{strat} , crudely mimics the stratosphere and is assumed to be transparent to solar (short wave) radiation and *optically thin* in the infra-red; its thermal (i.e. infra-red or long wave) absorptance is taken as $\varepsilon \ll 1$.⁷ By Kirchhoff's law (see [Section 3.1.1](#)) the stratospheric infra-red emittance is also ε ; its infra-red transmittance is $1 - \varepsilon$. The ground, at temperature T_g , is assumed to emit as a black body, and the whole system is assumed to be in radiative equilibrium.

As in [Section 1.3.1](#) the unreflected incoming solar irradiance F_0 is defined by [equation \(1.3\)](#) and we shall again take its value as 240 W m^{-2} . A useful related quantity is the **effective emitting temperature** of the Earth, defined by

$$T_e \equiv \left(\frac{F_0}{\sigma} \right)^{1/4} \approx 255 \text{ K.} \quad (3.36)$$

This is the temperature (assumed uniform) of the planet in the absence of an absorbing atmosphere, given the observed unreflected solar irradiance F_0 : see [Section 1.3.2](#).

⁷ In fact on a global and annual average the stratosphere absorbs about 12 W m^{-2} of solar radiation and has a mean infra-red absorptance of about 0.1.

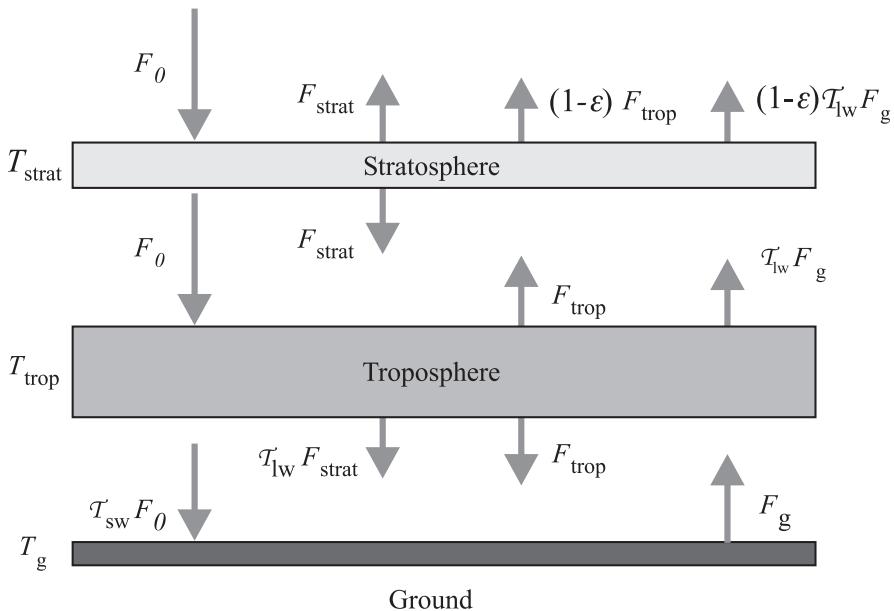


Fig. 3.20 Two-layer model atmosphere, including an optically thin stratosphere at temperature T_{strat} , a troposphere at temperature T_{trop} and the ground at temperature T_g . See text for further details.

The model is illustrated in Figure 3.20; balancing the upward and downward irradiances above the stratosphere we have

$$F_0 = F_{\text{strat}} + (1 - \varepsilon) (F_{\text{trop}} + \mathcal{T}_{\text{lw}} F_g). \quad (3.37)$$

Here

$$F_{\text{strat}} = \sigma \varepsilon T_{\text{strat}}^4, \quad F_{\text{trop}} = \sigma (1 - \mathcal{T}_{\text{lw}}) T_{\text{trop}}^4, \quad F_g = \sigma T_g^4,$$

are the upward emission from the stratosphere, the troposphere and the ground, respectively, and the effects of the non-zero absorptances of each atmospheric layer have been taken into account.

The balance of irradiances between the stratosphere and the troposphere implies

$$F_0 + F_{\text{strat}} = F_{\text{trop}} + \mathcal{T}_{\text{lw}} F_g; \quad (3.38)$$

eliminating F_0 between equations (3.37) and (3.38) therefore gives

$$2F_{\text{strat}} = \varepsilon (F_{\text{trop}} + \mathcal{T}_{\text{lw}} F_g). \quad (3.39)$$

This has a simple physical interpretation: the left-hand side represents the net radiative power leaving the stratosphere in the upward and downward directions; in equilibrium this must equal the net power absorbed by the stratosphere from the troposphere and the ground, as represented by the right-hand side. The solar irradiance does not appear in this equation, since it has been assumed to pass through the stratosphere without absorption.

Eliminating $F_{\text{trop}} + T_{\text{lw}}F_g$ from equations (3.37) and (3.38) we obtain

$$F_0 - F_{\text{strat}} = (1 - \varepsilon)(F_0 + F_{\text{strat}})$$

and hence

$$\sigma \varepsilon T_{\text{strat}}^4 = F_{\text{strat}} = \frac{\varepsilon F_0}{2 - \varepsilon}. \quad (3.40)$$

But since $\varepsilon \ll 1$ we obtain to a good approximation

$$\sigma T_{\text{strat}}^4 \approx \frac{F_0}{2},$$

independently of ε , and using equation (3.36) this gives the stratospheric temperature

$$T_{\text{strat}} \approx \frac{T_e}{2^{1/4}} = 214 \text{ K}. \quad (3.41)$$

This value does not depend on the absorptance ε of the stratosphere, provided that it is small, nor on the power $F_{\text{trop}} + T_{\text{lw}}F_g$ coming from below, provided that it is non-zero; it depends only on the unreflected solar irradiance F_0 .

From equations (3.39) and (3.40) we obtain

$$F_{\text{trop}} = \frac{2F_0}{2 - \varepsilon} - T_{\text{lw}}F_g. \quad (3.42)$$

The balance of irradiances between the troposphere and the ground implies

$$T_{\text{sw}}F_0 + T_{\text{lw}}F_{\text{strat}} + F_{\text{trop}} = F_g; \quad (3.43)$$

using equations (3.40), (3.42) and (3.43) we can obtain expressions for F_{trop} and F_g in terms of F_0 . These are similar to equations (1.6) and (1.5), but include small increases of order ε , due to the ‘greenhouse effect’ of the optically thin stratosphere. The resulting tropospheric and surface temperatures are therefore slightly greater than in the absence of the stratosphere.

3.7.2 Continuously stratified atmosphere in radiative equilibrium

We now consider an atmosphere in radiative equilibrium in which the temperature varies continuously with height. Our aim is to solve the radiative-transfer equations, subject to appropriate boundary conditions, to find the vertical temperature profile. In the infra-red we use the diffuse approximation of Section 3.2.3 and further assume that the atmosphere is **grey**: that is, the extinction coefficient k (and hence the scaled optical depth χ^* defined by equations (3.29) and (3.16)) is independent of frequency.⁸ We can therefore integrate equation (3.15) and the corresponding equation for upward irradiance over frequency to

⁸ This ‘grey gas’ assumption is convenient for some purposes, but it can be seriously misleading for others: see e.g. Section 8.5. Note that Figure 3.14 implies that the transmittance – which depends on k – varies strongly with frequency for the most important infra-red-absorbing gases.

obtain the following equations for the spectrally integrated long-wave irradiances $F^\uparrow(\chi^*)$ and $F^\downarrow(\chi^*)$:

$$-\frac{dF^\uparrow}{d\chi^*} + F^\uparrow = \pi B(T), \quad (3.44a)$$

$$\frac{dF^\downarrow}{d\chi^*} + F^\downarrow = \pi B(T). \quad (3.44b)$$

Equations (3.44) represent the **two-stream approximation**; note that B is the spectrally integrated Planck function, so that $\pi B(T) = \sigma T^4$ by equation (3.7), where T is the temperature at the level corresponding to the scaled optical depth χ^* . In each of equations (3.44) the first term represents the rate of change of the irradiance along the path, the second term represents extinction and the term πB on the right-hand side represents emission, which contributes equally in both directions; cf. Section 3.2.2. The minus sign in equation (3.44a) appears because χ^* is an optical depth and decreases along the upward path.

We again assume that the atmosphere is transparent to short-wave radiation, so there can be no short-wave heating ($Q^{sw} = 0$). Given that the atmosphere is in radiative equilibrium the long-wave heating Q^{lw} is also zero; see Section 3.6.4. Hence, by equation (3.28), the net upward long-wave irradiance F_z is independent of height, so

$$F_z = F^\uparrow - F^\downarrow = \text{constant}.$$

The constant can be found by considering the boundary condition at the top of the atmosphere. Here the scaled optical depth $\chi^* = 0$ and the downward long-wave irradiance $F^\downarrow(0) = 0$ also, since the only incoming radiation is short-wave. The net upward long-wave irradiance here is therefore $F_z = F^\uparrow(0)$, which must balance the incoming unreflected short-wave irradiance, which we again take as $F_0 \approx 240 \text{ W m}^{-2}$, as in Section 1.3.2. Hence

$$F_z = F^\uparrow - F^\downarrow = F_0. \quad (3.45)$$

We now find F^\uparrow , F^\downarrow and πB by a series of tricks. We first add equations (3.44a) and (3.44b) to get

$$-\frac{d}{d\chi^*} (F^\uparrow - F^\downarrow) + F^\uparrow + F^\downarrow = 2\pi B(T);$$

the derivative vanishes by equation (3.45), so

$$\pi B(T) = \frac{1}{2} (F^\uparrow + F^\downarrow). \quad (3.46)$$

We next subtract equation (3.44b) from equation (3.44a) and use equation (3.45) to get

$$\frac{d}{d\chi^*} (F^\uparrow + F^\downarrow) = F^\uparrow - F^\downarrow = F_0.$$

Since F_0 is constant, we can integrate to get

$$F^\uparrow + F^\downarrow = F_0 \chi^* + \text{constant}.$$

Again using the upper boundary condition, we see that the constant is F_0 , so that

$$F^\uparrow + F^\downarrow = F_0(1 + \chi^*). \quad (3.47)$$

From equations (3.45) and (3.47) we can then find the upward and downward irradiances in terms of F_0 and the scaled optical depth,

$$F^\uparrow = \frac{1}{2}F_0(2 + \chi^*), \quad (3.48)$$

$$F^\downarrow = \frac{1}{2}F_0\chi^*, \quad (3.49)$$

and hence, using equation (3.46), an expression for the temperature:

$$\pi B(T) = \sigma T^4 = \frac{1}{2}F_0(1 + \chi^*). \quad (3.50)$$

Thus F^\uparrow , F^\downarrow and B are all linear functions of the scaled optical depth χ^* .

The expressions derived above apply within the atmosphere; we now consider the radiation balance at the ground, which we take to be a black body at temperature T_g and at scaled optical depth χ_g^* , say. From equation (3.48) the upward long-wave irradiance in the atmosphere just above the ground is $F_0(2 + \chi_g^*)/2$. If we assume that this equals the black-body irradiance $\pi B(T_g) = \sigma T_g^4$ from the ground itself, then

$$\sigma T_g^4 = F_0(1 + \frac{1}{2}\chi_g^*) = \sigma T_e^4(1 + \frac{1}{2}\chi_g^*), \quad (3.51)$$

where $T_e \approx 255$ K is the effective emitting temperature given by equation (3.36). Note that this again demonstrates a greenhouse effect, since the atmosphere has a non-zero optical depth, $\chi_g^* > 0$, and so T_g is larger than the value T_e that would occur in the absence of an absorbing atmosphere.

The scaled optical depth is not a very intuitive vertical coordinate. However, if the height profiles of the extinction coefficient k and the absorbing gas density ρ_a are given, we can express the radiances and temperature in terms of height. As a simple example, we take k to be independent of height z and $\rho_a(z) = \rho_a(0) e^{-z/H_a}$, as in Section 3.6.2. Then $\chi^*(z) = \chi_g^* e^{-z/H_a}$, and equations (3.48)–(3.50) give

$$F^\uparrow(z) = \frac{1}{2}F_0 \left(2 + \chi_g^* e^{-z/H_a} \right), \quad F^\downarrow(z) = \frac{1}{2}F_0 \chi_g^* e^{-z/H_a},$$

and

$$T(z) = \left[\frac{F_0}{2\sigma} \left(1 + \chi_g^* e^{-z/H_a} \right) \right]^{1/4}. \quad (3.52)$$

These profiles are plotted as functions of z/H_a in Figure 3.21 for $\chi_g^* = 2$ and $F_0 = 240 \text{ W m}^{-2}$. The height at which the temperature T equals the effective emitting temperature T_e (see equation (3.36)) is sometimes called the **emission height**; in the present case, equation (3.52) shows that this is the height at which the scaled optical depth $\chi^* = 1$. Moreover $T \rightarrow (F_0/2\sigma)^{1/4}$ as $z \rightarrow \infty$: this is sometimes called the **skin temperature**, and it equals the stratospheric temperature $T_{\text{strat}} = 2^{-1/4}T_e$ given by equation (3.41) in the simpler model of Section 3.7.1. The behaviour of the atmospheric temperature as the ground is approached is given by

$$T(z) \rightarrow T_b \equiv T_e \left(\frac{1 + \chi_g^*}{2} \right)^{1/4} \quad \text{as } z \downarrow 0,$$

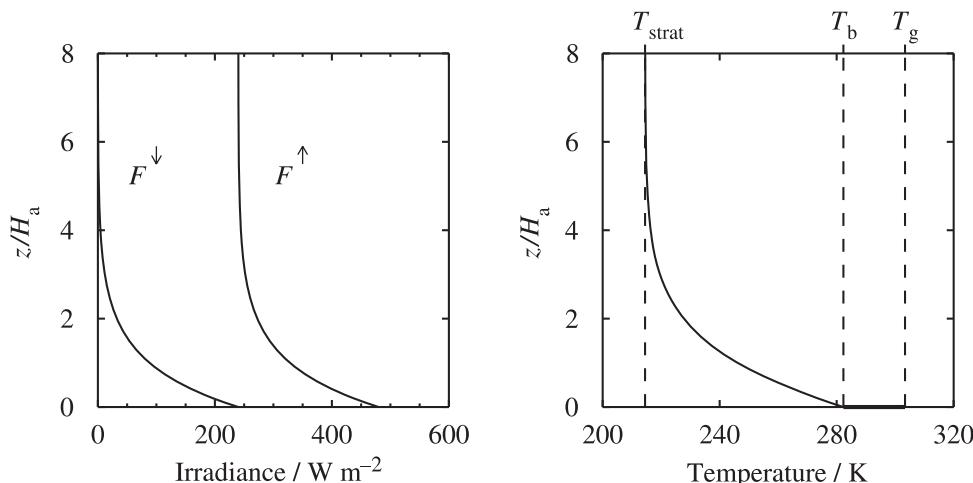


Fig. 3.21 Vertical profiles of the downward and upward irradiances (left) and temperature (right) for the two-stream model. Here $x_g^* = 2$ and $F_0 = 240 \text{ W m}^{-2}$. In the right-hand panel the vertical dashed lines indicate the temperatures T_{strat} , T_b and T_g (which equal 215 K, 282 K and 303 K, respectively, for the stated parameters), and the thick horizontal line at $z = 0$ indicates the temperature discontinuity between the bottom of the atmosphere and the ground.

whereas the temperature T_g of the ground itself is given by equation (3.51) as

$$T_g \equiv T_e \left(\frac{2 + x_g^*}{2} \right)^{1/4}.$$

The model therefore predicts a temperature discontinuity between the ground and the atmosphere just above it: this is of course an unphysical consequence of the radiative-equilibrium assumption. Moreover the lapse rate $-dT/dz$ associated with the radiative-equilibrium profile (3.52) may exceed the dry adiabatic lapse rate near the ground, implying static instability there. Inclusion of convection in the model removes the temperature discontinuity and forces the lapse rate to become stable.

3.8 A simple model of scattering

We conclude this chapter with a brief discussion of the scattering of solar radiation in the atmosphere. We consider a plane-parallel atmosphere and make the two-stream approximation, assuming that a fraction f of the spectral irradiance is scattered forwards (i.e. in the direction of the beam) and a fraction $1 - f$ is scattered backwards. These fractions represent integrals over the forward- and backward-pointing hemispheres of the corresponding radiances.

We first work in terms of the vertical coordinate z and consider the downward penetration of the downward spectral irradiance F_v^\downarrow at a short-wave frequency v . By analogy with the

radiative-transfer equation (3.10), this satisfies the equation

$$\frac{dF_v^\downarrow}{d(-z)} = -a_v \rho_a F_v^\downarrow - s_v \rho_a (1-f) F_v^\downarrow + s_v \rho_a (1-f) F_v^\uparrow. \quad (3.53)$$

This states that the downward irradiance decreases in the downward direction because of the absorption term $a_v \rho_a F_v^\downarrow$ and because of the backscattered component $s_v \rho_a (1-f) F_v^\downarrow$, and increases because of the backscattering of upward irradiance $s_v \rho_a (1-f) F_v^\uparrow$ into the downward direction. Emission is neglected, since the Planck function for short-wave frequencies at terrestrial atmospheric temperatures is negligible. A rearrangement of equation (3.53), in terms of the extinction coefficient $k_v = a_v + s_v$ (see equation (3.9)) gives

$$\frac{dF_v^\downarrow}{d(-z)} = -k_v \rho_a F_v^\downarrow + s_v \rho_a f F_v^\downarrow + s_v \rho_a (1-f) F_v^\uparrow.$$

We now replace $-k_v \rho_a dz$ by the optical depth increment $d\chi_v^*$; cf. equation (3.29). Introducing the **albedo for single scattering**,

$$\omega_v = \frac{s_v}{a_v + s_v} = \frac{s_v}{k_v},$$

and dropping the subscripts v for simplicity, we obtain

$$\frac{dF^\downarrow}{d\chi^*} + F^\downarrow = \omega [f F^\downarrow + (1-f) F^\uparrow]. \quad (3.54a)$$

In the same way we obtain the equation for the upward irradiance F^\uparrow :

$$-\frac{dF^\uparrow}{d\chi^*} + F^\uparrow = \omega [f F^\uparrow + (1-f) F^\downarrow]. \quad (3.54b)$$

The simplest case is that of an atmospheric layer in which there is scattering, but no absorption, i.e. with $\omega = 1$. This layer could for example be a crude representation of a cloud. Equations (3.54) then give

$$\frac{dF^\downarrow}{d\chi^*} + (1-f)(F^\downarrow - F^\uparrow) = 0, \quad (3.55a)$$

$$-\frac{dF^\uparrow}{d\chi^*} + (1-f)(F^\uparrow - F^\downarrow) = 0. \quad (3.55b)$$

Addition of equations (3.55) shows that the net downward spectral irradiance is constant:

$$-F_z = F^\downarrow - F^\uparrow = \text{constant}. \quad (3.56)$$

Suppose that the downward spectral irradiance at frequency v at the top of the layer ($\chi^* = 0$, say) is F_{inc} and that the upward spectral irradiance at the bottom of the layer ($\chi^* = \chi_1^*$, say) is zero. Suppose also that the transmitted spectral irradiance at the bottom is F_{trans} and the reflected spectral irradiance at the top is F_{refl} ; see Figure 3.22. We therefore have the boundary conditions

$$F^\downarrow(0) = F_{\text{inc}}, \quad F^\uparrow(0) = F_{\text{refl}}, \quad F^\downarrow(\chi_1^*) = F_{\text{trans}}, \quad F^\uparrow(\chi_1^*) = 0. \quad (3.57)$$

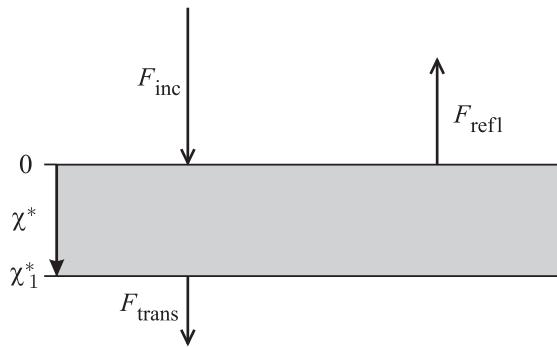


Fig. 3.22 Scattering from a horizontal layer, showing the incoming spectral irradiance F_{inc} , the transmitted spectral irradiance F_{trans} and the reflected spectral irradiance F_{refl} .

We can define a transmission coefficient $T = F_{\text{trans}}/F_{\text{inc}}$ and a reflection coefficient (or layer albedo) $R = F_{\text{refl}}/F_{\text{inc}}$. Then $-F_z = F_{\text{inc}} - F_{\text{refl}} = (1 - R)F_{\text{inc}} = F_{\text{trans}} - 0 = Tf_{\text{inc}}$. Hence

$$R + T = 1, \quad (3.58)$$

implying conservation of irradiance.

Using $F^\downarrow - F^\uparrow = -F_z = Tf_{\text{inc}}$ and the upper boundary condition $F^\downarrow(0) = F_{\text{inc}}$, we can now solve equation (3.55a) to get

$$F^\downarrow(\chi^*) = F_{\text{inc}}[1 - (1 - f)T\chi^*] \quad (3.59a)$$

and, using equation (3.56),

$$F^\uparrow(\chi^*) = F_{\text{inc}}[R - (1 - f)T\chi^*]. \quad (3.59b)$$

Then substitution of $F^\uparrow(\chi_1^*) = 0$ into equation (3.59b) and use of equation (3.58) gives the transmission and reflection coefficients

$$T = \frac{1}{1 + (1 - f)\chi_1^*}, \quad R = \frac{(1 - f)\chi_1^*}{1 + (1 - f)\chi_1^*}.$$

Note that, as $\chi_1^* \rightarrow \infty$, the reflection coefficient (layer albedo) tends to 1 and the transmission coefficient tends to 0, i.e. the layer becomes a nearly perfect reflector as its optical thickness becomes large.

The case of a layer that both absorbs and scatters solar radiation, so that the single-scattering albedo $\omega < 1$, involves more complicated mathematics, but the physical conclusions are similar, except that $R + T < 1$. See Problem 3.11.

Further reading

For general introductions to the subject of atmospheric radiation see, for example, Houghton (2002), Wallace and Hobbs (2006), Salby (1996) and Goody (1995). A more specialised

text is that of Petty (2004). The work by Goody and Yung (1989) is the second edition of the classic text in the field, and goes far beyond the treatment given in this chapter. The proof that the spectral energy density in an isothermal cavity depends only on the frequency and temperature is given in books on thermal physics, such as that by Blundell and Blundell (2009). The Planck function, equation (3.1), and the Boltzmann distribution, equation (3.4), must be derived using statistical mechanics: see, for example, the books by Blundell and Blundell (2009) and Glazer and Wark (2001). For the Poynting vector see standard books on electromagnetism, such as those by Grant and Phillips (1990) and Lorrain *et al.* (1988). The integrating factor method is given in standard texts on mathematical methods, such as Riley *et al.* (2006), Boas (1983) and Lyons (1995). Banwell and McCash (1994) and Atkins (2006) provide introductions to molecular spectroscopy, while an excellent treatment of many aspects of radiative transfer and spectroscopy is presented by Thorne (1988). Derivations of the Lorentz profile, equation (3.21), are given, for example, by Thorne (1988) and Goody and Yung (1989); see also Bransden and Joachain (1989) for the quantum-mechanical details. The quantum harmonic oscillator is covered, for example, by Rae (2007). The cooling-to-space approximation was introduced by Rodgers and Walshaw (1966). Basic radiative properties of the stratosphere are given by Ramanathan and Dickinson (1979).

Problems

Problem 3.1 (i) Given equation (3.1) for B_ν , derive equation (3.2) for B_λ .

(ii) Show that the values of ν and λ that maximise $B_\nu(T)$ and $B_\lambda(T)$ are given by

$$\frac{\nu_{\max}}{T} = c_1, \quad \lambda_{\max} T = c_2,$$

respectively, where c_1 and c_2 are constants. Show that, for any given T , ν_{\max} and λ_{\max} do not correspond to the same photon energy. Given that $c_2 = 2.9 \times 10^{-3} \text{ m K}$, find the temperatures for which B_λ is maximum at 500 nm and at 10 μm .

(iii) Assuming that the Sun behaves as a black body at a temperature of $T = 5800 \text{ K}$, calculate $B_\nu(T)$ and $B_\lambda(T)$ at 500 nm and the percentage increase in these spectral radiances if the temperature increases by 100 K.

Problem 3.2 (i) Elements of surface ΔS_1 and ΔS_2 are a distance r apart, with normals inclined at angles θ_1 and θ_2 to the line joining them. Under what conditions is the net radiative power flow between them in the optical passband $\Delta\nu$ given by

$$\Delta P_{1 \rightarrow 2} = \frac{B_\nu(T_1)}{r^2} \Delta\nu \Delta S_1 \Delta S_2 \cos\theta_1 \cos\theta_2?$$

(ii) Integrate $\int_0^\infty B_\nu d\nu$ using the substitution $x = h\nu/(k_B T)$ and hence derive the right-hand member of equation (3.7). Use the relations

$$\int_0^\infty \frac{x^3 dx}{e^x - 1} = \frac{\pi^4}{15}, \quad \sigma = \frac{2\pi^5 k^4}{15h^3 c^2}.$$

Hence use the results of (i) to show that the total solar irradiance F_s (i.e. the normal irradiance just outside the atmosphere) equals $\sigma T_s^4 r_s^2/d^2$, where T_s is the effective black-body temperature of the Sun, r_s is the Sun's radius and d is the Sun–Earth distance. Evaluate F_s , given that $T_s = 5800$ K.

(iii) By carrying out the appropriate integration, find the power absorbed by a horizontal black disc of radius a from an infinite horizontal black plane at temperature T_1 over which it is suspended at height h .

(iv) Find the answer to (iii) using symmetry and thermodynamic arguments. Find also the power absorbed from the plane by a black sphere of radius a .

Problem 3.3 Assume that the Earth and Sun radiate as black bodies at 288 and 6000 K, respectively. Show that the radiative power per unit area, per unit wavelength interval (the spectral irradiance), falling on a horizontal plane just above the Earth's surface and with the Sun overhead, is $\pi B_\lambda(288\text{ K})$ from the side facing the Earth and

$$(1 - A) \left(\frac{r_s}{d}\right)^2 \pi B_\lambda(6000\text{ K})$$

from the side facing the Sun. Here $A \approx 0.3$ is the albedo of the Earth–atmosphere system (see [Section 1.3.1](#)), r_s is the Sun's radius and d is the Sun–Earth distance. Use [Figure 3.1](#) to estimate the maximum values of the solar and terrestrial spectral irradiances and hence sketch the variations of these irradiances with wavelength. Comment on the practical relevance of your results.

Problem 3.4 A low-altitude Earth satellite in an equatorial orbit carries below it two isolated spherical radiation sensors of negligible heat capacity. One is painted white and may be assumed to be perfectly reflecting at all wavelengths at which there is significant solar energy ($\lambda < 4\text{ }\mu\text{m}$) and perfectly absorbing at longer wavelengths. The other is painted black and is perfectly absorbing at all wavelengths. Assume that there is no input of direct or scattered energy from the spacecraft.

(i) Calculate the temperatures of the two spheres at midnight over a thick cloud sheet at a temperature of 280 K.

(ii) Find the radiance due to diffusely scattered sunlight just above a cloud of albedo α when the Sun is overhead.

(iii) Assuming that the spheres are shadowed from overhead direct sunlight by the spacecraft, calculate their temperatures over a thick cloud of temperature 270 K and albedo 0.8, given that the total solar irradiance $F_s = 1370\text{ W m}^{-2}$.

Problem 3.5 Investigate the equivalent width of a Lorentz line, as follows. Substitute S times the right-hand side of [equation \(3.21\)](#) into [equation \(3.25\)](#) and use [equation \(3.26\)](#) to get the following expression for the equivalent width:

$$\mathcal{W} = \gamma_L \int_{-\infty}^{\infty} \left[1 - \exp\left(-\frac{q}{x^2 + 1}\right) \right] dx \quad (3.60)$$

where $x = (\nu - \nu_0)/\gamma_L$ and $q = u_a S / (\gamma_L \pi)$.

The integral in [equation \(3.60\)](#) can be evaluated in terms of Bessel functions; however, there are two special cases in which it can be approximated easily. The first is the case of a **strong line**, which is valid when the mass of absorber in the path is so large that $u_a S / (\gamma_L \pi) \gg 1$, i.e. $q \gg 1$: sketch the integrand in [equation \(3.60\)](#) in this case. Verify that the exponential in the integrand is then negligible except where $x^2 \gtrsim q \gg 1$ and show that the equivalent width becomes

$$\mathcal{W} \approx 2\gamma_L \int_0^\infty [1 - \exp(-q/x^2)] dx.$$

Use the substitution $y = 1/x$ and integration by parts to show that

$$\mathcal{W} \approx 4\gamma_L q \int_0^\infty \exp(-qy^2) dy = 2\gamma_L (\pi q)^{1/2} = 2(Su_a \gamma_L)^{1/2}.$$

On the other hand, if the mass of absorber in the path is small, we have the **weak-line** limit, with $q \ll 1$. Show that, in this case,

$$\mathcal{W} \approx \gamma_L \int_{-\infty}^\infty \frac{q dx}{x^2 + 1} = \gamma_L q \pi = Su_a,$$

in agreement with [equation \(3.27\)](#).

Problem 3.6 Under what conditions on the absorber mixing ratio is the transmittance at the centre ν_0 of a Lorentz line, for a path at pressure p , of length l and absorber density ρ_a , independent of the pressure?

Problem 3.7 Investigate the equivalent width of a Doppler line, as follows. Put $r = u_a S / \gamma_D \sqrt{\pi}$ and $x = (\nu - \nu_0) / \gamma_D$, to get

$$\mathcal{W} = \gamma_D \int_{-\infty}^\infty \left[1 - \exp(-re^{-x^2}) \right] dx.$$

The integral cannot be evaluated explicitly in general. However, in the *strong-line* limit, $r \gg 1$, it can be shown that the absorptance $\mathcal{A}_\nu = 1 - \mathcal{T}_\nu$ (the integrand [...] above) is nearly zero for $|x| > (\ln r)^{1/2}$ and nearly unity for $|x| < (\ln r)^{1/2}$. Hence show that the integral can be evaluated approximately as

$$\mathcal{W} = 2\gamma_D \int_0^\infty (1 - \mathcal{T}_\nu) dx \approx 2\gamma_D (\ln r)^{1/2} = 2\gamma_D \left[\ln \left(\frac{u_a S}{\gamma_D \sqrt{\pi}} \right) \right]^{1/2}.$$

Show that, in the *weak-line* limit $r \ll 1$,

$$\mathcal{W} \approx \gamma_D \int_{-\infty}^\infty r \exp(-x^2) dx = \gamma_D r \sqrt{\pi} = Su_a,$$

in agreement with [equation \(3.27\)](#).

Problem 3.8 Consider a CO₂ line at 15 μm (wavenumber $\tilde{\nu} = 1/\lambda = 667 \text{ cm}^{-1}$) with $\gamma_L = (p/p_0)(T_0/T)^{1/2} \gamma_{L0}$ (cf. [equation \(3.22\)](#)), where $p_0 = 1000 \text{ hPa}$, $T_0 = 250 \text{ K}$ and $\gamma_{L0} = 3 \times 10^9 \text{ Hz}$. Find the approximate pressure level in the atmosphere at which the

transmittance of a horizontal path at the line centre under the assumption of pure Doppler broadening is the same as that under the assumption of pure Lorentz broadening. (Assume that the atmosphere is at a typical temperature of 250 K.)

Problem 3.9 The Curtis–Godson approximation is a method of calculating approximate transmittances for paths of varying pressure and absorber density. For a vertical path, one defines a mean pressure \bar{p} and mean temperature \bar{T} , weighted by the absorber density:

$$\bar{p} = \frac{1}{u_a} \int p \rho_a dz, \quad \bar{T} = \frac{1}{u_a} \int T \rho_a dz,$$

where $u_a = \int \rho_a dz$ is the total mass per unit transverse cross-sectional area of the path.

Show that the spectral means of atmospheric transmittances calculated using the Curtis–Godson approximation are accurate when (i) strong pressure-broadening conditions exist or (ii) the absorption is weak, irrespective of the line shape.

Derive \bar{p} for an atmospheric absorber with a constant mass mixing ratio μ for a vertical path from height z to ∞ .

Problem 3.10 (i) Write down an expression for the heating rate h ($= Q/c_p$ in units of K s^{-1}) at height z in the atmosphere due to an absorber of density $\rho_a(z)$ and constant extinction coefficient k when the Sun is overhead. (Denote the normal solar spectral irradiance integrated over the absorption band by $F_{vs} \Delta\nu$; neglect scattering.)

(ii) A simple model of radiative heating by solar radiation in the upper stratosphere and lower mesosphere assumes that ozone has a mass mixing ratio

$$\mu(p) = ap^{1/2},$$

where a is a constant and p is pressure. Show that the heating rate h due to ozone is proportional to

$$p^{1/2} \exp\left(-\frac{2ak}{3g} p^{3/2}\right),$$

where k is the extinction coefficient of ozone for solar radiation. Show that the pressure p_m at which h is a maximum is close to that of the stratopause and evaluate $h(p_m)$ in K day^{-1} .

(Take $k = 1.5 \times 10^4 \text{ m}^2 \text{ kg}^{-1}$ in a spectral region where the solar irradiance is 7 W m^{-2} and take $a = 3 \times 10^{-7} \text{ Pa}^{-1/2}$.)

(iii) In a simple model of radiative cooling due to thermal emission by CO_2 ('cooling to space'; see [Section 3.6.3](#)), the cooling rate C is given (in units of K s^{-1}) by

$$C = -\frac{g}{c_p} \left(\frac{dT^*}{dp} \right) \pi B(T) \Delta\nu,$$

where $T^*(p)$ is the transmittance from pressure level p to space (cf. [equation \(3.34\)](#)). The spectrally integrated Planck function $B(T) \Delta\nu$ for the CO_2 band can be approximated by

$$B(T) \Delta\nu = 15(T/300)^{3.8} \text{ W m}^{-2} \text{ steradian}^{-1}$$

and \mathcal{T}^* can be taken to have the form

$$\mathcal{T}^* = \exp(-\beta p)$$

where $\beta = 3.6 \times 10^{-4} \text{ Pa}^{-1}$. Diurnal variations can be crudely accounted for by dividing $h(p_m)$ by 2 (why?). Estimate the radiative equilibrium temperature at p_m .

Problem 3.11 Consider a layer of cloud that both absorbs and scatters solar radiation, so that the single-scattering albedo $\omega < 1$. Put $1 - \omega f = a$ and $\omega(1 - f) = b$ and rewrite equations (3.54) as

$$\left(\frac{d}{d\chi^*} + a \right) F^\downarrow = b F^\uparrow, \quad \left(\frac{d}{d\chi^*} - a \right) F^\uparrow = -b F^\downarrow.$$

Hence show that

$$\left(\frac{d^2}{d\chi^{*2}} - \alpha^2 \right) F^\uparrow = 0, \quad \text{where} \quad \alpha^2 = a^2 - b^2.$$

Verify that $\alpha^2 > 0$, so the solution for F^\uparrow is a combination of exponentials or, equivalently, of hyperbolic functions. Using one of the boundary conditions (3.57), show that F^\uparrow must take the form

$$F^\uparrow(\chi^*) \propto \sinh \alpha (\chi_1^* - \chi^*).$$

Show that F^\downarrow involves both sinh and cosh functions and, by applying the remaining boundary conditions, derive the reflection coefficient R and the transmission coefficient T for the cloud:

$$R = \frac{b \sinh \theta}{\alpha \cosh \theta + a \sinh \theta}, \quad T = \frac{\alpha}{\alpha \cosh \theta + a \sinh \theta},$$

where $\theta = \alpha \chi_1^*$.

Show that, as the optical depth χ_1^* of the layer becomes large,

$$R \rightarrow R_\infty = \frac{b}{\alpha + a} = \frac{b}{(a^2 - b^2)^{1/2} + a}, \quad T \rightarrow 0.$$

Evaluate the reflection and transmission coefficients for $\omega = 0.9997$, $f = 0.9$ and $\chi_1^* = 10, 20$ and 30 . Note that $R + T < 1$: what is the physical reason for this? Find also the reflection coefficient for an infinitely deep cloud with these values of ω and f .

A wide variety of fluid flows occurs in the atmosphere. This chapter introduces the basic fluid-dynamical laws that govern these atmospheric flows. The length scales of interest range from metres to thousands of kilometres; these are many orders of magnitude greater than molecular scales such as the mean free path, at least in the lower and middle atmosphere. We may therefore average over many molecules, ignoring individual molecular motions and regarding the fluid as continuous. ‘Local’ values of quantities such as density, temperature and velocity may be defined at length scales that are much greater than the mean free path but much less than the scales on which the meteorological motion varies.

In [Section 4.1](#) we derive the mass conservation law (often called the **continuity equation**) for a fluid. In [Section 4.2](#) we introduce the concept of the material derivative and the Eulerian and Lagrangian views of fluid motion. An alternative form of the mass conservation law is given in [Section 4.3](#) and the equation of state for the atmosphere (an ideal gas) is recalled in [Section 4.4](#). Then in [Section 4.5](#) Newton’s Second Law is applied to a continuous fluid, giving the **Navier–Stokes equation**. The Earth’s rotation cannot be ignored for large-scale atmospheric flows, so its incorporation into the Navier–Stokes equation is discussed in [Section 4.6](#). The full equations of motion for a spherical Earth and for Cartesian tangent-plane geometry are given in [Section 4.7](#). Simplifications of these equations for large-scale flows are introduced in [Section 4.8](#). A useful alternative formulation with pressure, rather than height, as a vertical coordinate is briefly mentioned in [Section 4.9](#). Finally, in [Section 4.10](#), we introduce the First Law of Thermodynamics, as applied to the atmosphere. Some applications of these results to atmospheric phenomena are discussed in the problems at the end of the chapter, and further applications are provided in [Chapter 5](#).

4.1 Mass conservation

Consider a small ‘box’, or volume element, of sides Δx , Δy and Δz , fixed in space, with fluid passing through it ([Figure 4.1](#)). Let the fluid velocity be $\mathbf{u} = (u, v, w)$ and the fluid density be ρ ; these generally vary with position \mathbf{r} and time t .

The net mass inflow in the x -direction per unit time is the difference between that coming in on the left and that going out on the right (see [Figure 4.2](#)); i.e.

$$\begin{aligned} & \left[\rho(x)u(x) - \rho(x + \Delta x)u(x + \Delta x) \right] \Delta y \Delta z \\ & \approx -\frac{\partial}{\partial x}(\rho u) \Delta x \cdot \Delta y \Delta z = -\frac{\partial}{\partial x}(\rho u) \Delta V, \end{aligned}$$

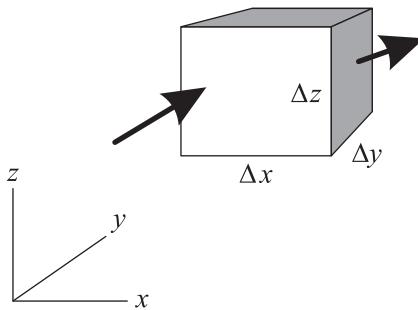


Fig. 4.1

A volume element, fixed in space, with fluid passing into and out of it.

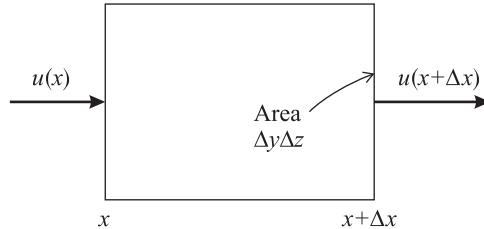


Fig. 4.2

A side view of the box in Figure 4.1.

where $\Delta V = \Delta x \Delta y \Delta z$, the volume of the box. Including the inflow and outflow in the y - and z -directions, the net inflow is therefore

$$-\nabla \cdot (\rho \mathbf{u}) \Delta V. \quad (4.1)$$

This must equal the rate of increase of mass (density times volume) in the box, namely

$$\frac{\partial}{\partial t}(\rho \Delta V). \quad (4.2)$$

However, the box is fixed, so its volume ΔV is constant. Hence, on cancelling ΔV from equations (4.1) and (4.2), we obtain the **continuity equation**, or **mass conservation law**

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0. \quad (4.3)$$

By expansion of the $\nabla \cdot (\rho \mathbf{u})$ term using a standard vector-calculus identity, this can also be written as

$$\frac{\partial \rho}{\partial t} + \mathbf{u} \cdot \nabla \rho + \rho \nabla \cdot \mathbf{u} = 0. \quad (4.4)$$

4.2 The material derivative

We can measure flow with respect to fixed points \mathbf{r} in space. For example, an anemometer measures wind speed and a wind-vane measures wind direction; together these give us the

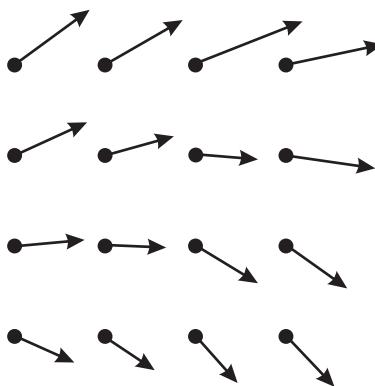


Fig. 4.3

A schematic diagram of fluid flow vectors at each point on a grid.

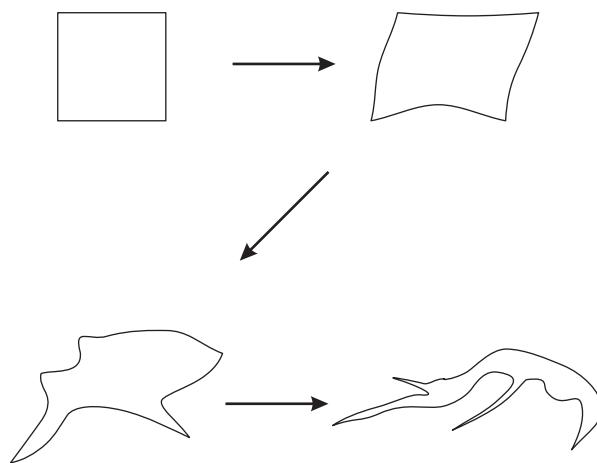


Fig. 4.4

A schematic illustration of the possible distortion with time of a moving blob of fluid, initially rectangular in shape.

vector wind \mathbf{u} . Suppose that such measurements are made for a fixed grid of points, as shown schematically in Figure 4.3; this is the **Eulerian** picture of the flow.

An alternative is to imagine following the motion of tiny ‘blobs’ of fluid (e.g. marked by a dye). Each blob is supposed to contain many molecules, but moves as a coherent entity, at least for short times; however, in general a blob will be distorted by the flow, as shown in Figure 4.4.

Let the position of a given blob (or rather, of its centre of mass) at time t be $\mathbf{r}(t)$. The trajectory of its motion between times 0 and t might be as in Figure 4.5. Now consider the kinematics of the blob, ignoring any distortion. Its velocity at time t is the Eulerian velocity \mathbf{u} evaluated at the current position $\mathbf{r}(t)$ of the blob; this is also equal to the current time rate-of-change of the blob’s position. Hence

$$\mathbf{u}(\mathbf{r}(t), t) = \frac{d\mathbf{r}}{dt}. \quad (4.5)$$

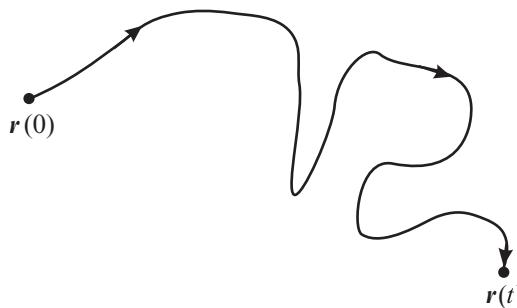


Fig. 4.5

A schematic picture of the trajectory of a moving blob of fluid. On this scale, the blob is too small for its distortion in shape to be shown.

In a similar fashion the acceleration of the blob at time t , say $\mathbf{a}(\mathbf{r}(t), t)$, equals the second derivative of the position vector:

$$\mathbf{a}(\mathbf{r}(t), t) = \frac{d^2\mathbf{r}}{dt^2}.$$

This picture of fluid motion in terms of moving blobs is called the **Lagrangian** picture. In practice we may not be able to follow blobs for long; nevertheless, the Lagrangian picture is important conceptually.

The connection between the Lagrangian and Eulerian pictures is made as follows. Let $\mathbf{r} = (x, y, z)$ and $\mathbf{u} = (u, v, w)$. The x -component of equation (4.5) states that

$$\frac{dx}{dt} = u(x(t), y(t), z(t), t),$$

so, differentiating with respect to t once again and using the chain rule on the right,

$$\begin{aligned} \frac{d^2x}{dt^2} &= \frac{\partial u}{\partial x} \frac{dx}{dt} + \frac{\partial u}{\partial y} \frac{dy}{dt} + \frac{\partial u}{\partial z} \frac{dz}{dt} + \frac{\partial u}{\partial t} \\ &= \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} \\ &= \frac{\partial u}{\partial t} + \mathbf{u} \cdot \nabla u \equiv \frac{D\mathbf{u}}{Dt}, \end{aligned}$$

(where the three components of equation (4.5) have been used) or in vector form

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2} = \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = \frac{D\mathbf{u}}{Dt}. \quad (4.6)$$

We have introduced here the important operator

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla, \quad (4.7)$$

known as the **material derivative** or **advection derivative**. This represents the rate of change with respect to time *following the motion* (or following a blob) and should be contrasted to $\partial/\partial t$, the rate of change with respect to time *at a fixed point*. (The notation d/dt is also sometimes used for the material derivative in meteorological books.)

Note that the contribution $(\mathbf{u} \cdot \nabla) \mathbf{u}$ to the material derivative is **non-linear** in \mathbf{u} . This makes the behaviour of the atmosphere difficult to forecast (for example, non-linear systems are well known to display chaotic behaviour), but also leads to many interesting features.

4.3 An alternative form of the continuity equation

Using the definition (4.7), the continuity equation (4.4) can be rewritten as

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{u} = 0. \quad (4.8)$$

This form of the continuity equation has a useful interpretation in Lagrangian terms, as follows.

Rather than considering a fixed ‘box’ in space, as above, we focus on a small, moving, initially rectangular ‘blob’.¹ The blob is of fixed mass δm and time-varying volume $\delta V = \delta x \delta y \delta z$ and travels in a unidirectional flow $(u(x), 0, 0)$ in the x -direction; see Figure 4.6. This flow is assumed to vary in the x -direction; hence, within a small time interval δt the left-hand end of the blob moves from x to $x_1 = x + u(x) \delta t$ while the right-hand end moves from $x + \delta x$ to $x_2 = x + \delta x + u(x + \delta x) \delta t$. The blob’s x -length therefore increases by an amount $[u(x + \delta x) - u(x)] \delta t$. Since the velocity vector has no y - or z -component, the dimensions δy and δz remain unchanged. The new volume of the blob is therefore

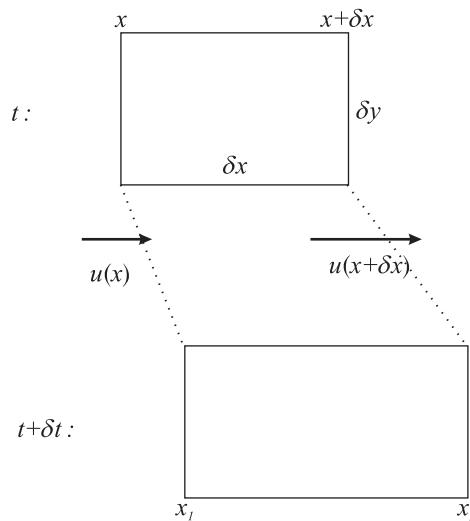


Fig. 4.6 A rectangular blob, at times t and $t + \delta t$, in a flow that is moving in the x -direction and also varies in the x -direction.

¹ In this chapter we use δ to represent small Lagrangian quantities and Δ to represent small Eulerian quantities. We also use the word ‘box’ to indicate an infinitesimal volume element *fixed in space* and ‘blob’ to indicate a *moving* infinitesimal mass element. The latter is similar to the ‘parcel’ concept introduced in Section 2.5.

$\delta V' = \{\delta x + [u(x + \delta x) - u(x)]\delta t\} \delta y \delta z$ and this may be approximated, using a Taylor expansion, by

$$\delta V' \approx \left(\delta x + \frac{\partial u}{\partial x} \delta x \delta t \right) \delta y \delta z = \delta V \left(1 + \frac{\partial u}{\partial x} \delta t \right).$$

So

$$\frac{D(\delta V)}{Dt} = \lim_{\delta t \rightarrow 0} \left(\frac{\delta V' - \delta V}{\delta t} \right) = (\delta V) \frac{\partial u}{\partial x}.$$

More generally, allowing for the flow to be three-dimensional and varying in all directions, so that $\mathbf{u} = \mathbf{u}(\mathbf{r}, t)$, we find that

$$\frac{D(\delta V)}{Dt} = (\delta V) \nabla \cdot \mathbf{u}$$

However, $\delta V = \delta m / \rho$ and the constant mass δm can be cancelled, giving

$$\frac{D\rho}{Dt} = -\rho \nabla \cdot \mathbf{u},$$

which is equivalent to [equation \(4.8\)](#). If $\nabla \cdot \mathbf{u} = 0$ everywhere, the fluid is said to be **incompressible**: see [Section 5.2](#).

4.4 The equation of state for the atmosphere

As in [Chapter 2](#), we assume that the atmosphere is an ideal gas, so that the equation of state is

$$p = R_a T \rho, \quad (4.9)$$

where p is the pressure, T is the temperature, ρ is the density and R_a is the gas constant per unit mass of air; cf. [equation \(2.2\)](#).

4.5 The Navier–Stokes equation

In this section we derive the **Navier–Stokes equation** for a fluid, by applying Newton's Second Law to a small moving blob of fluid. Since the blob is moving, this means that we are using the Lagrangian perspective.

Assume that the blob is instantaneously of cuboidal shape, with sides δx , δy and δz , as in [Figure 4.7](#). The volume of the blob is $\delta V = \delta x \delta y \delta z$ and its mass is $\rho \delta V$, where the density is ρ , as above. Newton's Second Law states that

$$(\rho \delta V) \mathbf{a} = \delta \mathbf{F}, \quad (4.10)$$

where \mathbf{a} is the acceleration of the blob and $\delta \mathbf{F}$ is the vector sum of the forces (including pressure forces, gravity and friction) acting on the blob. We must consider each of these forces in turn.

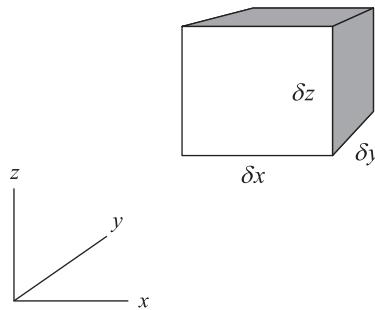


Fig. 4.7

An instantaneous view of a small blob of fluid.

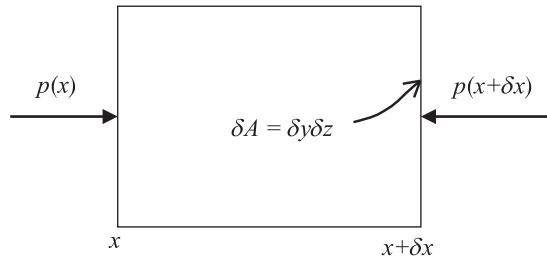


Fig. 4.8

Illustrating the various pressure forces acting on the blob in the x -direction.

First, consider the pressure forces acting on the blob and single out the x -direction (see Figure 4.8). At position x the pressure force is $p(x) \delta A$ in the positive x -direction, where $\delta A = \delta y \delta z$ is the area of the relevant wall of the blob; at position $x + \delta x$ the pressure force is

$$p(x + \delta x) \delta A \approx \left(p(x) + \frac{\partial p}{\partial x} \delta x + \dots \right) \delta A,$$

in the negative x -direction, using a Taylor expansion again. The net pressure force in the positive x -direction is therefore

$$-\delta x \frac{\partial p}{\partial x} \delta A = -\delta V \frac{\partial p}{\partial x},$$

so in three dimensions the vector pressure force is

$$\delta \mathbf{F}_{\text{press}} = -(\delta V)(\nabla p). \quad (4.11)$$

Next consider the gravity force acting on the blob: this is just the mass $\rho \delta V$ of the blob times g , acting downwards:

$$\delta \mathbf{F}_{\text{grav}} = -(\rho \delta V)g \mathbf{k}, \quad (4.12)$$

where \mathbf{k} is the unit vertical vector, $(0, 0, 1)$.

Finally, consider the viscous forces acting on the blob. These require quite a complicated treatment in general, so we shall consider a special case by way of illustration. Suppose

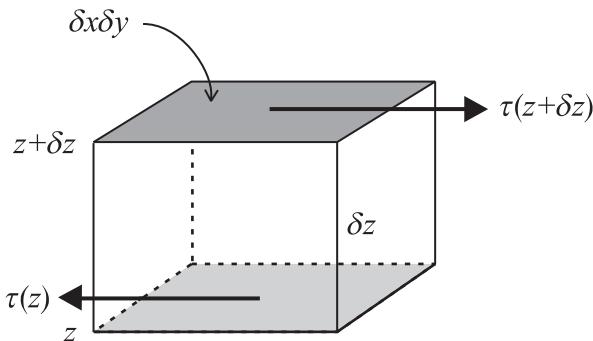


Fig. 4.9

Illustrating the horizontal stresses acting on the top and bottom of the blob.

first that the bulk flow (averaged over many molecules) is in the x -direction only, and varies only in the z -direction, i.e. $\mathbf{u} = (u(z), 0, 0)$. Recall, from the kinetic theory of gases, that the x -component of the viscous stress (or tangential force per unit area) acting on a horizontal surface from above is given by

$$\tau = \eta \frac{du}{dz}, \quad (4.13)$$

where η is the **dynamic viscosity**. (It is usually called μ in fluid dynamics texts.) An equal and opposite stress acts on the surface from below. Figure 4.9 illustrates how stresses in opposite directions and of slightly different magnitude act on the top and bottom surfaces of our blob. The net contribution to the viscous force on the blob is the difference between the stresses on the top and bottom, namely

$$[\tau(z + \delta z) - \tau(z)]\delta x \delta y \approx \frac{\partial \tau}{\partial z} \delta x \delta y \delta z = \frac{\partial \tau}{\partial z} \delta V; \quad (4.14)$$

using equation (4.13) and assuming that η is constant, this becomes

$$\eta \frac{d^2 u}{dz^2} \delta V. \quad (4.15)$$

The situation is more complicated if there are y - and z -variations as well, but it turns out that the viscous vector force is

$$\delta \mathbf{F}_{\text{visc}} = \delta V \eta \left(\nabla^2 \mathbf{u} + \frac{1}{3} \nabla (\nabla \cdot \mathbf{u}) \right). \quad (4.16)$$

This is clearly a generalisation of equation (4.15); note that it simplifies for an incompressible fluid, for which $\nabla \cdot \mathbf{u} = 0$.

Assuming that these are the only relevant forces acting on the blob, putting expressions (4.11), (4.12) and (4.16) into Newton's Second Law (4.10) and using equation (4.6) for the acceleration \mathbf{a} , we get

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \nabla p - g\mathbf{k} + \mathbf{F}_{\text{visc}}, \quad (4.17)$$

where $\mathbf{F}_{\text{visc}} = \delta \mathbf{F}_{\text{visc}} / \delta V$. This is the **Navier–Stokes equation** (or **momentum equation**) for compressible fluid flow in an inertial frame.

4.6 Rotating frames of reference

The rotation of the Earth is significant for the large-scale dynamics of the atmosphere. It is usually most convenient to work with a coordinate system fixed with respect to the Earth; since this system is rotating with respect to inertial space, modifications to the Navier–Stokes equation (4.17) must be made.

These modifications involve changes to the time derivatives of vectors. Suppose that the frame R rotates at a constant angular velocity Ω with respect to an inertial frame I and define the z-axes of both frames to be in the direction of Ω . Now consider a time-varying vector $A(t)$ and suppose, for simplicity, that it lies in the common x, y plane of R and I. As viewed in the rotating frame R, A changes from $A(t)$ to $A(t + \delta t) \approx A(t) + \delta A_R$, in the time interval between t and $t + \delta t$, where

$$\delta A_R \equiv \left(\frac{dA}{dt} \right)_R \delta t,$$

as illustrated in Figure 4.10. However, if the same vector A is viewed in the inertial frame I, we must allow for the rotation of the frame R with respect to the frame I in the time δt ; this gives an extra contribution to the change in A when it is viewed in the inertial frame, as shown in Figure 4.11. With a little consideration of vector geometry, the extra contribution to the change in A is found to be $\Omega \times A \delta t$, for small δt . (The same can be shown to hold if A is not in the x, y plane.) So, as $\delta t \rightarrow 0$, we obtain the following relationship between time derivatives in the inertial frame I and the rotating frame R:

$$\left(\frac{dA}{dt} \right)_I = \left(\frac{dA}{dt} \right)_R + \Omega \times A. \quad (4.18)$$

A double application of equation (4.18) gives

$$\left(\frac{d^2A}{dt^2} \right)_I = \left(\frac{d^2A}{dt^2} \right)_R + 2\Omega \times \left(\frac{dA}{dt} \right)_R + \Omega \times (\Omega \times A).$$

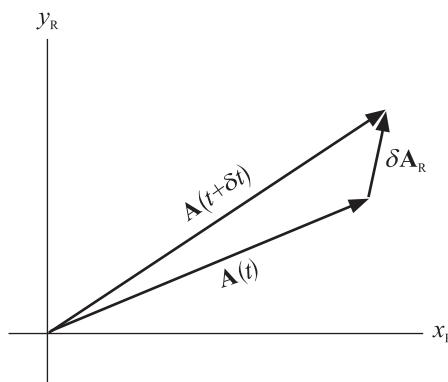


Fig. 4.10

The change of the vector A viewed in the rotating frame R.

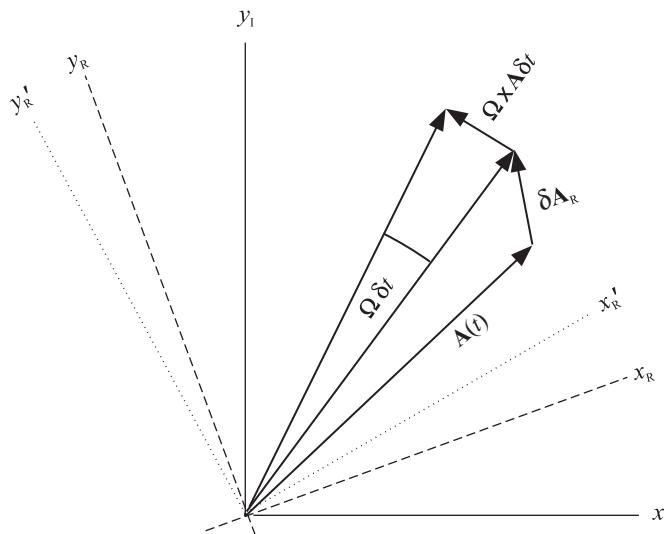


Fig. 4.11 The change of the vector \mathbf{A} viewed in the inertial frame I. In time δt the axes fixed in R, as observed in I, rotate through an angle $\Omega \delta t$, from $x_R y_R$ to $x'_R y'_R$.

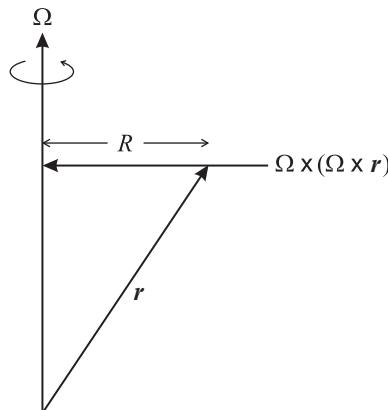


Fig. 4.12

Illustrating the centripetal acceleration.

In particular, if $\mathbf{A} = \mathbf{r}$, the position vector, we obtain the following relationship between the acceleration in the inertial frame and the acceleration, velocity and position vector in the rotating frame:

$$\mathbf{a}_I = \mathbf{a}_R + 2\Omega \times \mathbf{u}_R + \Omega \times (\Omega \times \mathbf{r}). \quad (4.19)$$

Let us examine the terms involving Ω on the right-hand side of equation (4.19). The term $2\Omega \times \mathbf{u}_R$ represents the **Coriolis acceleration**: it is perpendicular both to the velocity \mathbf{u}_R in the rotating frame and to Ω . The term $\Omega \times (\Omega \times \mathbf{r})$ represents the **centripetal acceleration**: it has magnitude $\Omega^2 R$, where R is the perpendicular distance from the point \mathbf{r} to the rotation axis, and is directed perpendicularly towards the rotation axis, as shown in Figure 4.12.

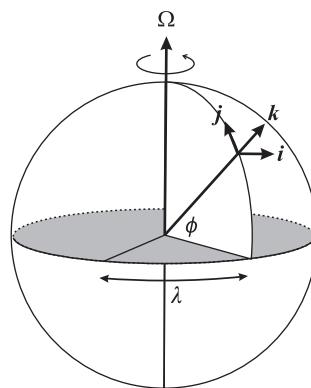


Fig. 4.13

Local Cartesian coordinate axes on a spherical Earth.

To rewrite the Navier–Stokes equation in the rotating frame, we need to replace the acceleration $\mathbf{a} = D\mathbf{u}/Dt$ by \mathbf{a}_l in [equation \(4.17\)](#), to get the corresponding equation *with respect to the rotating frame*:

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \nabla p - 2\boldsymbol{\Omega} \times \mathbf{u} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) - g\mathbf{k} + \mathbf{F}_{\text{visc}}, \quad (4.20)$$

where the subscript R has been dropped and \mathbf{u} now represents the velocity measured in the rotating frame R. Here the acceleration terms $2\boldsymbol{\Omega} \times \mathbf{u}$ and $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r})$ are written with minus signs on the right-hand side of [equation \(4.20\)](#) and can be regarded as **fictitious forces** (per unit mass) in Newton's Second Law in the rotating frame. The term $-2\boldsymbol{\Omega} \times \mathbf{u}$ is known as the **Coriolis force** and the term $-\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r})$ as the **centrifugal force**.² The gravity and centrifugal forces can be combined to give the **effective gravity** $\mathbf{g}' = -g\mathbf{k} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r})$, although the difference between \mathbf{g}' and $-g\mathbf{k}$ is small at the Earth's surface (see [Problem 4.5](#)).

4.7 Equations of motion in coordinate form

4.7.1 Spherical coordinates

The natural coordinates in which to express our equations, when they are applied to the Earth, are spherical coordinates (r, ϕ, λ) ,³ where r is the distance from the centre of the Earth, ϕ is latitude and λ is longitude. At a point on the Earth's surface we draw unit vectors \mathbf{i} pointing eastwards, \mathbf{j} pointing northwards and \mathbf{k} pointing upwards, as shown in [Figure 4.13](#).

² Note that the acceleration is centripetal ('centre-seeking') whereas the fictitious force, being of opposite sign, is centrifugal ('centre-fleeing').

³ Strictly speaking, oblate spheroidal coordinates should be used; see [Gill \(1982\)](#). We shall ignore this complication here.

It is convenient to introduce small incremental distances $dx = r \cos \phi d\lambda$ in the eastward (or **zonal**) direction and $dy = r d\phi$ in the northward (or **meridional**) direction. We also introduce the vertical distance z from the Earth's surface, so that $r = a + z$, where a is the Earth's radius and $dz = dr$.

It is shown in [Appendix B](#) that, neglecting the centripetal acceleration, equation (4.20) can be written in component form as follows:

$$\frac{Du}{Dt} - \left(2\Omega + \frac{u}{r \cos \phi} \right) (v \sin \phi - w \cos \phi) + \frac{1}{\rho} \frac{\partial p}{\partial x} = F^{(x)}, \quad (4.21a)$$

$$\frac{Dv}{Dt} + \frac{wv}{r} + \left(2\Omega + \frac{u}{r \cos \phi} \right) u \sin \phi + \frac{1}{\rho} \frac{\partial p}{\partial y} = F^{(y)}, \quad (4.21b)$$

$$\frac{Dw}{Dt} - \frac{u^2 + v^2}{r} - 2\Omega u \cos \phi + \frac{1}{\rho} \frac{\partial p}{\partial z} + g = F^{(z)}, \quad (4.21c)$$

where $F^{(x)}$, $F^{(y)}$ and $F^{(z)}$ are components of the frictional force in the eastward, northward and upward directions, respectively, and

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + w \frac{\partial}{\partial z}. \quad (4.22)$$

4.7.2 Approximations to the spherical equations

Equations (4.21) are complicated, but approximate versions are sufficient for modelling many atmospheric dynamical phenomena. In the first place, we can replace the distance r by the Earth's radius a with negligible error, since the part of the atmosphere in which we are interested has a depth of 100 km or so, which is much less than $a \approx 6400$ km.

Having done this, consider equation (4.21a). Two useful simplifications can be made here.

- The zonal wind will generally be less than 100 m s^{-1} in magnitude. It can then be verified that

$$\frac{|u|}{a \cos \phi} \ll 2\Omega,$$

except perhaps near the poles, where $\cos \phi \rightarrow 0$. (In this calculation we can use the fact that Ωa , the tangential speed of the Earth's surface at the equator, is approximately 465 m s^{-1} .)

- Vertical velocities are usually much less than horizontal velocities, so

$$|w \cos \phi| \ll |v \sin \phi|,$$

except perhaps near the equator, where $\sin \phi \rightarrow 0$.

Given these two results, and introducing the **Coriolis parameter**,

$$f = 2\Omega \sin \phi,$$

equation (4.21a) reduces to

$$\frac{Du}{Dt} - fv + \frac{1}{\rho} \frac{\partial p}{\partial x} = F^{(x)}. \quad (4.23a)$$

Equation (4.21b) can be simplified in a similar way (but using also the result that $|wv|/r \ll 2\Omega|u \sin \phi|$, except possibly near the equator), to give

$$\frac{Dv}{Dt} + fu + \frac{1}{\rho} \frac{\partial p}{\partial y} = F^{(y)}. \quad (4.23b)$$

In equation (4.21c) it is easy to show that the terms $(u^2 + v^2)/a$ and $2\Omega u \cos \phi$ are very much smaller than $g = 9.8 \text{ m s}^{-2}$ for any reasonable values of the horizontal velocity components u and v . Omitting also the vertical friction term $F^{(z)}$, which is usually regarded as negligible, we obtain

$$\frac{Dw}{Dt} + \frac{1}{\rho} \frac{\partial p}{\partial z} + g = 0. \quad (4.23c)$$

4.7.3 Tangent-plane geometry

Equations (4.23a)–(4.23c) are expressed in spherical coordinates. However, the use of dx and dy as small eastward and northward distances suggests a further useful simplification, which is valid when we are considering a comparatively small region near a point P at latitude ϕ_0 and longitude λ_0 . In this case we can introduce Cartesian coordinates (x, y, z) on the **tangent plane** at the point P, as shown in Figure 4.14: clearly there is little difference between distances $(x', y'$, say) on the surface of the Earth and distances (x, y) on the tangent plane in the neighbourhood of P. We can therefore re-interpret equations (4.23a)–(4.23c) as applying to the Cartesian coordinates on the tangent plane and avoid complications due to spherical geometry. We must, however, replace f (which varies with latitude) by the

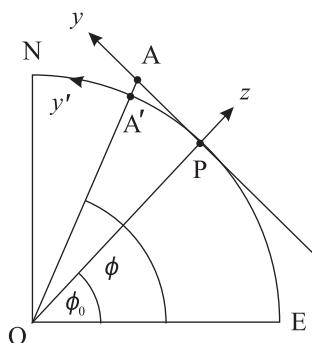


Fig. 4.14

Illustrating the use of tangent-plane geometry. The figure shows a section through the centre of the Earth, O, the North Pole, N, and the point P (at latitude ϕ_0); E denotes a point on the equator. The plane AP is tangential to the Earth at point P. Northward distances y are measured in this plane, whereas northward distances y' are measured on the surface of the Earth. These distances are almost equal if the latitude ϕ of point A' on the surface of the Earth is close to ϕ_0 .

constant value

$$f_0 = 2\Omega \sin \phi_0.$$

Note that equations (4.23a) and (4.23b) then become identical to those for a system that is rotating about the z -axis with angular velocity $\frac{1}{2}f_0$, except that centripetal accelerations are again neglected. This approximation is called **the f -plane approximation** and the analogous system is **the f -plane**.

If we wish to consider a larger region we may retain the tangent-plane approximation but allow for some variation of f with latitude. A Taylor expansion of $f(\phi)$ about $\phi = \phi_0$ gives

$$f(\phi) = 2\Omega \sin \phi = 2\Omega[\sin \phi_0 + (\phi - \phi_0) \cos \phi_0 + \dots]$$

so that on the tangent plane, where $\phi - \phi_0 \approx y/a$,

$$f(y) \approx f_0 + \beta y, \quad \text{where} \quad \beta = \frac{2\Omega \cos \phi_0}{a} = \left(\frac{df}{dy} \right)_{y=0}. \quad (4.24)$$

Equation (4.24) is called **the β -plane approximation**: instead of taking f to be constant, as in the f -plane approximation, we allow it to vary linearly with the northward distance y .

4.8 Geostrophic and hydrostatic approximations

Under appropriate dynamical conditions we can simplify equations (4.23a)–(4.23c) still further, using the method of **scale analysis**. (A simple form of scale analysis was used in Section 4.7.2 to simplify equations (4.21a)–(4.21c).) For example, consider motions associated with **synoptic-scale** systems – that is, large-scale weather systems – at midlatitudes, with the time and space scales given in Table 4.1. First consider typical sizes of the terms in equation (4.23c), the vertical momentum equation. The material derivative can be

Table 4.1 Some scales for large-scale motion in the atmosphere.

Scale	Symbol	Typical magnitude
Horizontal scale	L	$1000 \text{ km} = 10^6 \text{ m}$
Vertical scale	H	$10 \text{ km} = 10^4 \text{ m}$
Horizontal velocity	U	10 m s^{-1}
Vertical velocity	W	10^{-2} m s^{-1}
Timescale	T	$1 \text{ day} \sim 10^5 \text{ s}$
Surface density	ρ	1 kg m^{-3}
Earth's radius	a	$6.4 \times 10^6 \text{ m}$
$2 \times$ rotation rate	2Ω	10^{-4} s^{-1}
Acceleration of gravity	g	10 m s^{-2}

expanded as

$$\frac{Dw}{Dt} = \frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z}.$$

Very roughly, we can estimate the individual terms here as

$$\frac{\partial w}{\partial t} \sim \frac{W}{T} \sim \frac{10^{-2}}{10^5} \sim 10^{-7} \text{ m s}^{-2},$$

$$u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} \sim \frac{UW}{L} \sim \frac{10^{-1}}{10^6} \sim 10^{-7} \text{ m s}^{-2},$$

$$w \frac{\partial w}{\partial z} \sim \frac{W^2}{H} \sim \frac{10^{-4}}{10^4} \sim 10^{-8} \text{ m s}^{-2};$$

hence in total we estimate

$$\frac{Dw}{Dt} \sim 10^{-7} \text{ m s}^{-2}.$$

This is very much smaller than the g term in [equation \(4.23c\)](#):

$$g \sim 10^1 \text{ m s}^{-2}.$$

The remaining term is $(1/\rho) \partial p / \partial z$; this is the only term that can possibly balance the large g term, so [equation \(4.21c\)](#) must become, to a good approximation,

$$\frac{\partial p}{\partial z} = -g\rho. \quad (4.25)$$

This shows that, under our assumed scaling, the vertical momentum equation reduces to **hydrostatic balance** (cf. [Chapter 2, equation \(2.12\)](#)).

Now let us perform a similar scale analysis on the horizontal momentum [equation \(4.23a\)](#):

$$\frac{\partial u}{\partial t} \sim 10^{-4}, \quad u \frac{\partial u}{\partial x} \sim 10^{-4}, \quad w \frac{\partial u}{\partial z} \sim 10^{-5},$$

$$fv = 2\Omega v \sin \phi \sim 10^{-3}$$

(all in units of m s^{-2}), where a midlatitude value of ϕ is used and $F^{(x)}$ is assumed negligible.

Here fv is the biggest term (just), so it must be balanced by the remaining term $(1/\rho) \partial p / \partial x$, giving the **geostrophic approximation**:

$$fv = \frac{1}{\rho} \frac{\partial p}{\partial x}; \quad (4.26a)$$

a similar scale analysis applied to [equation \(4.23b\)](#) yields

$$-fu = \frac{1}{\rho} \frac{\partial p}{\partial y}. \quad (4.26b)$$

Thus, for synoptic-scale systems, the horizontal momentum [equations \(4.23a\)](#) and [\(4.23b\)](#) reduce to **geostrophic balance**, in which the horizontal pressure gradients are balanced by Coriolis forces associated with the horizontal winds. The following points should be noted.

- At each height the horizontal wind blows along the **isobars** (the lines of constant pressure), since $(u, v, 0) \cdot \nabla p = 0$ under the geostrophic approximation.
- It is twice the *vertical* component of the rotation vector, $2\Omega \sin \phi$, that enters.
- The ratio of the ‘horizontal advection’ term to the ‘Coriolis term’ in equation (4.23a) is approximately

$$\frac{u \partial u / \partial x}{fv} \sim \frac{U^2 / L}{fU} = \frac{U}{fL} \equiv \text{Ro}; \quad (4.27)$$

Ro is a dimensionless number, called the **Rossby number**. If $\text{Ro} \gg 1$, the Coriolis term is comparatively small, whereas if $\text{Ro} \ll 1$, the Coriolis term is comparatively large and the geostrophic approximation is usually valid.

4.8.1 The thermal windshear equations

From the geostrophic approximation (4.26a) and the ideal gas law (4.9) we have

$$fv = \frac{R_a T}{p} \frac{\partial p}{\partial x} = R_a T \frac{\partial \ln p}{\partial x},$$

while from equations (4.25) and (4.9) we have

$$-\frac{g}{R_a T} = \frac{\partial \ln p}{\partial z}.$$

Neglecting vertical (but not horizontal) variations in T ,⁴ cross-differentiation of these gives

$$f \frac{\partial v}{\partial z} \approx \frac{g}{T} \frac{\partial T}{\partial x} \quad (4.28a)$$

and similarly, from equations (4.26b), (4.25) and (4.9), we obtain

$$f \frac{\partial u}{\partial z} \approx -\frac{g}{T} \frac{\partial T}{\partial y}. \quad (4.28b)$$

These are called the **thermal wind equations** or, more correctly, the **thermal windshear equations**; they give a very useful relation between horizontal temperature gradients and vertical gradients of the horizontal wind, when both geostrophic balance and hydrostatic balance apply.

As an example of the use of the thermal windshear equations, consider a case in which the temperature varies only with the northward distance y , so $T = T(y)$, and decreases with y , so $dT/dy < 0$. Suppose that the horizontal wind direction varies between $\mathbf{u}_1 = (u_1, v_1)$ at a lower level $z = z_1$ and $\mathbf{u}_2 = (u_2, v_2)$ at a higher level $z = z_2$, as shown in Figure 4.15. Since T is independent of x , the isotherms $T = \text{constant}$ are parallel to the x -axis; moreover, equation (4.28a) then implies that $\partial v / \partial z = 0$, so $v_2 = v_1$. Therefore the difference between the wind vectors at z_1 and z_2 is parallel to the x -axis and hence to the isotherms $T = \text{constant}$. Given that $dT/dy < 0$, equation (4.28b) implies that $\partial u / \partial z > 0$, so $u_2 > u_1$.

⁴ Vertical variations of T need not be neglected if we use pressure p , rather than height z , as a vertical coordinate; see Section 4.9.

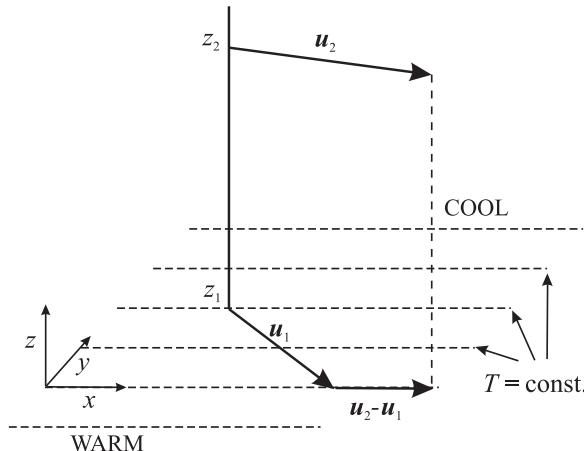


Fig. 4.15

Illustrating thermal windshear balance.

Consider the Northern Hemisphere, where $f > 0$: it can then be seen that if the wind turns anticlockwise with height (known as **backing**), the wind on average blows across the isotherms from the cold to the warm side: this is called **cold advection**. Conversely, if the wind turns clockwise with height (**veering**), the wind on average blows across the isotherms from warm to cold (**warm advection**).

Another example is the zonally averaged (i.e. longitudinally averaged) zonal (i.e. east–west) winds in the lower and middle atmosphere, shown in Figure 1.6, which tend to be nearly in thermal-windshear balance with the zonally averaged temperature, shown in Figure 1.5. It can readily be checked that the signs, at least, of $\partial u / \partial z$ and $\partial T / \partial y$ in these figures are consistent with equation (4.28b).

There are many other applications of geostrophic and hydrostatic balance in the interpretation of atmospheric processes. An important case is the structure of fronts; see Problem 4.8.

4.8.2 A circular vortex: gradient-wind balance

We now consider another important type of approximation to the horizontal momentum equations, which is more accurate than geostrophic balance. In fact it is exact for purely horizontal flow in the form of a steady vortex with circular streamlines. We consider such a vortex, centred at a point P, near latitude ϕ , and use polar coordinates r and φ , in the horizontal plane, also centred at P, as shown in Figure 4.16. For this circular vortex the velocity vector is $\mathbf{u} = V(r) \mathbf{i}_\varphi$, where \mathbf{i}_φ is the unit vector in the azimuthal direction, and the pressure p is a function of r only. The corresponding Coriolis force is $-\rho f \mathbf{k} \times \mathbf{u} = \rho f V \mathbf{i}_r$, where \mathbf{i}_r is the unit vector in the radial direction. We also include the centrifugal force $(\rho V^2/r) \mathbf{i}_r$. The sum of these two forces has to be balanced by a radial pressure gradient, so

$$\frac{V^2}{r} + fV = \frac{1}{\rho} \frac{dp}{dr} \equiv G,$$

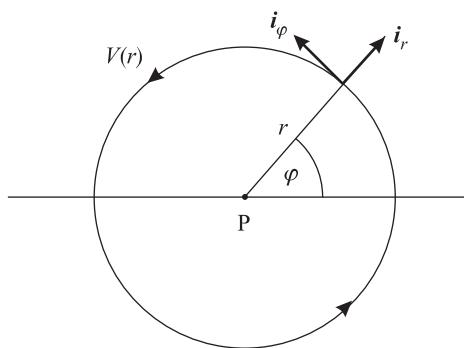


Fig. 4.16

Illustrating the circular vortex.

say. This equation expresses **gradient-wind balance** and should be contrasted with geostrophic balance, in which the V^2/r term is neglected. It is a quadratic equation for V , if G is given, with solutions

$$V(r) = -\frac{rf}{2} \pm \left(\frac{r^2 f^2}{4} + Gr \right)^{1/2}.$$

Note that physically sensible solutions are possible only if V is real, i.e. if

$$G = \frac{1}{\rho} \frac{dp}{dr} \geq -rf^2/4,$$

so the pressure cannot drop too rapidly with radius if gradient-wind balance is to occur.

4.9 Pressure coordinates and geopotential

Meteorologists often use pressure p , rather than height z , as a vertical coordinate. This has two, quite separate, advantages: first, atmospheric measurements are usually referenced with respect to pressure rather than height; and second, as we shall see, the equations of motion take a simpler form in these ‘pressure coordinates’ than they do in height coordinates. However, care needs to be exercised when taking partial derivatives: we need to be clear about which variables are held constant in the partial differentiation!

For a small column of air, $dp = -g\rho dz$ from hydrostatic balance (equation (4.25)), so the pressure decreases monotonically with height. There is thus a one-to-one relationship between pressure and height, which means that no ambiguity will arise in using p as a vertical coordinate.

We can turn equation (4.25) upside down and use the ideal gas law, equation (4.9), to get

$$g \left(\frac{\partial z}{\partial p} \right)_{x,y,t} = - \left(\frac{1}{\rho} \right) = - \frac{R_a T}{p},$$

so, if the **geopotential** $\Phi = gz$ is introduced, the hydrostatic equation becomes

$$\frac{\partial \Phi}{\partial p} = -\frac{R_a T}{p} \quad (4.29)$$

in pressure coordinates. In equation (4.29) the subscripts x, y and t (indicating which variables are held constant when the partial p -derivative of Φ is taken) are omitted, since there should be no danger of confusion.

We can derive the pressure-coordinate versions of geostrophic balance as follows. We first eliminate the density ρ from equations (4.25) and (4.26a):

$$fv = -g \left(\frac{\partial p}{\partial x} \right) / \left(\frac{\partial p}{\partial z} \right) = -g \left(\frac{\partial p}{\partial x} \right) \left(\frac{\partial z}{\partial p} \right),$$

where we have again turned $(\partial p / \partial z)_{x,y,t}$ upside down. We next use the well-known ‘reciprocity’ theorem of partial differentiation (used, for example, in thermodynamics):

$$\left(\frac{\partial \alpha}{\partial \beta} \right)_\gamma \left(\frac{\partial \beta}{\partial \gamma} \right)_\alpha \left(\frac{\partial \gamma}{\partial \alpha} \right)_\beta = -1. \quad (4.30)$$

This gives

$$fv = g \left(\frac{\partial z}{\partial x} \right)_{y,p,t} = \left(\frac{\partial \Phi}{\partial x} \right)_{y,p,t} \quad (4.31)$$

and similarly

$$fu = - \left(\frac{\partial \Phi}{\partial y} \right)_{x,p,t}. \quad (4.32)$$

Equations (4.31) and (4.32) are the equations for geostrophic balance in pressure coordinates. A p -coordinate version of the thermal windshear equations can be derived, without assuming that T is independent of z ; see Problem 4.9.

The full equations of motion (4.21a)–(4.21c) can be derived in pressure coordinates, but these will not be needed here. Note, however, that instead of the geometric vertical velocity w we must introduce a vertical ‘pressure velocity’

$$\omega \equiv \frac{Dp}{Dt},$$

which can be shown to be approximately equal to $-\rho gw$.

In pressure coordinates the equation of continuity of mass, equation (4.4) or (4.8), is replaced by the simpler form

$$\left(\frac{\partial u}{\partial x} \right)_{y,p,t} + \left(\frac{\partial v}{\partial y} \right)_{x,p,t} + \left(\frac{\partial \omega}{\partial p} \right)_{x,y,t} = 0, \quad (4.33)$$

with no time derivative or density ρ appearing. This can be obtained by considering a small moving blob of fluid, as in Figure 4.7, with volume $\delta V = \delta x \delta y \delta z$ and mass $\delta m = \rho \delta V = -\delta x \delta y \delta p / g$ (using the hydrostatic equation in the form $\delta p = -g \rho \delta z$). This mass is constant, following the motion of the blob, so $D(\delta m) / Dt = 0$ and hence

$$0 = \frac{1}{\delta m} \frac{D}{Dt} (\delta m) = \frac{g}{\delta x \delta y \delta p} \frac{D}{Dt} \left(\frac{\delta x \delta y \delta p}{g} \right).$$

Using the chain rule and the fact that $D(\delta x)/Dt = D(x_2 - x_1)/Dt = u_2 - u_1 = \delta u$, $D(\delta p)/Dt = D(p_2 - p_1)/Dt = \omega_2 - \omega_1 = \delta\omega$, etc., we get

$$\frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{\delta \omega}{\delta p} = 0,$$

which gives equation (4.33) as δx , δy and δp tend to zero.

4.10 The thermodynamic energy equation

The First Law of Thermodynamics for a moving blob of fluid of unit mass, undergoing small changes δS of entropy, δU of internal energy and δV of volume in time δt , can be written:

$$T \delta S = \delta U + p \delta V;$$

cf. Chapter 2, equation (2.17). As shown in equation (2.22), this can be written, for an ideal gas, as

$$T \delta S = c_p \delta T - \frac{1}{\rho} \delta p, \quad (4.34)$$

where c_p is the specific heat capacity at constant pressure.

Dividing equation (4.34) by δt , letting $\delta t \rightarrow 0$ and remembering that we are considering a moving (Lagrangian) mass, so that the time derivative is D/Dt , not $\partial/\partial t$, we then obtain

$$Q \equiv T \frac{DS}{Dt} = c_p \frac{DT}{Dt} - \frac{1}{\rho} \frac{Dp}{Dt}. \quad (4.35)$$

This is one form of the First Law of Thermodynamics, as used in atmospheric physics. (It is also commonly called the thermodynamic energy equation in meteorology.) The quantity Q , defined by the first member of equation (4.35), is called the **diabatic heating rate per unit mass**. For a blob in adiabatic (adiathermal and reversible) motion, $Q = 0$.⁵

The main physical processes contributing to Q in the lower and middle atmosphere are latent heating and cooling (from condensation and evaporation, respectively, of water vapour) and radiative heating and cooling (from absorption and emission of electromagnetic radiation; see Chapter 3). Note that, if $Q = 0$,

$$\frac{DT}{Dt} = \frac{1}{\rho c_p} \frac{Dp}{Dt}, \quad (4.36)$$

so, in the absence of diabatic heating, the temperature of a moving blob of air will increase if it descends (moves to higher pressure) and will decrease if it ascends. Another way

⁵ The use of Q here is quite common in atmospheric physics; however, under reversible conditions, it would be more consistent with the notation of Chapter 2, e.g. equation (2.19), to use $\lim(\delta Q/\delta t) = DQ/Dt$ for the diabatic heating rate.

of showing this is to note that, if the blob moves adiabatically, it conserves its **potential temperature**, θ :

$$\theta = T \left(\frac{p_0}{p} \right)^\kappa,$$

where $\kappa = R_a/c_p$; see equation (2.25). As p increases, T has to increase as well, to keep θ constant, which is consistent with equation (4.36).

It is frequently more convenient to work with potential temperature θ than with temperature T ; in the presence of diabatic heating, the thermodynamic energy equation (4.35), expressed in terms of θ , becomes

$$\frac{D\theta}{Dt} = \frac{Q}{c_p} \left(\frac{p}{p_0} \right)^{-\kappa}. \quad (4.37)$$

Further reading

Good introductory textbooks on fluid dynamics in general are those by Tritton (1988), Faber (1995) and Acheson (1990): the first two of these are specifically aimed at undergraduate physicists. The basics of atmospheric fluid dynamics, including a treatment of pressure coordinates, are covered well by Holton (2004). Marshall and Plumb (2008) provide a very nice introduction to the dynamics of the atmosphere and the ocean, with strong links to relevant laboratory experiments. Excellent advanced textbooks on atmospheric and oceanic fluid dynamics, going well beyond the scope of the present book, are those of Vallis (2006), Pedlosky (1987) and Gill (1982). For elementary kinetic theory of gases see, for example, Blundell and Blundell (2009). The ‘reciprocity’ theorem of partial differentiation, equation (4.30), is given, for example, by Blundell and Blundell (2009), Boas (1983) and Lyons (1995).

Problems

Problem 4.1 Starting from the ideal gas law in the form (2.6), estimate the number of molecules n/V of air per unit volume at ground level. Hence show that the mean separation between molecules there is a few nanometres. Estimate also the mean free path, given that it is of order $V/(\pi d^2 n)$, where d is an effective molecular diameter. Comment on the relevance of these estimates for the assumption that atmospheric flow may be described as the motion of a continuous fluid. At what altitude might this assumption start to break down?

Problem 4.2 Starting from the continuity equation in Eulerian form,

$$\partial\rho/\partial t + \nabla \cdot (\rho\mathbf{u}) = 0,$$

derive the Lagrangian form,

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{u} = 0,$$

by vector manipulations. Give a physical interpretation of the latter equation in terms of the rate of change of volume of a small moving ‘blob’ of fluid of fixed mass δm , by showing that the relative rate of expansion of the blob’s volume equals the divergence of the velocity field.

Problem 4.3 Consider planar motion. Suppose that the coordinates of a particle are (x, y) in an inertial frame F and (x', y') in a frame F' that rotates at a constant angular velocity Ω with respect to F . Show that

$$x' = x \cos(\Omega t) + y \sin(\Omega t), \quad y' = y \cos(\Omega t) - x \sin(\Omega t),$$

if $x' = x$ and $y' = y$ at $t = 0$.

Consider a particle that describes the following trajectory in F :

$$x(t) = R \cos(\Omega t) + a \sin(\Omega t), \quad y(t) = R \sin(\Omega t) + b \cos(\Omega t),$$

where R , a and b are positive constants. Show that the particle is undergoing two-dimensional simple harmonic motion. Its trajectory in F takes the form of an ellipse; draw a rough sketch of this ellipse, given that $a = 0.4R$ and $b = 0.8R$, indicating the particle’s direction of motion. (Hint: choose four appropriate values of Ωt .)

Now consider the particle’s trajectory, as viewed by an observer in the rotating frame F' . Show that in this frame the particle moves in a circle; find the radius and centre of this circle, and the particle’s angular velocity and direction of motion. By differentiating x' and y' twice with respect to t , show that the particle’s motion in F' is consistent with the effect of a Coriolis force. (This is an example of an **inertial oscillation**; see Durran (1993) for a detailed analysis.)

Problem 4.4 If a ball is thrown a horizontal distance of 100 m at 30° latitude in 4 s, what is its sideways deflection due to the Coriolis force?

Problem 4.5 Evaluate the dimensionless parameter $\Omega^2 a/g$ for the Earth. Show that, at the Earth’s surface, the magnitude $|\mathbf{g}'|$ of the effective gravity is about 0.7% less at the equator than at the poles, and the maximum angle between \mathbf{g}' and a vector pointing towards the centre of the Earth is about 0.1° .

Problem 4.6 Estimate the Rossby numbers for the following flows and comment on your results.

1. A hurricane at 20° N, with wind speed about 50 m s^{-1} and horizontal length scale about 100 km.
2. A tornado in the American Midwest, with wind speed about 100 m s^{-1} and horizontal length scale about 100 m.
3. Flow in a bathtub vortex.

Problem 4.7 Consider an atmospheric vortex in which the flow is steady, horizontal and independent of height, with circular streamlines, in a frame rotating with angular velocity Ω about the vertical. Neglecting friction, show that

$$\frac{u^2}{r} + 2\Omega u = \frac{1}{\rho} \frac{dp}{dr},$$

where r is the radial distance from the centre of the vortex, $u(r)$ is the (tangential) wind speed, $p(r)$ is the pressure and ρ is the density (assumed constant).

Hence explain why highs are regions of weak pressure gradients and gentle winds, but lows can have large pressure gradients and strong winds. Is the geostrophic wind an overestimate or underestimate of the wind in a low-pressure system?

Problem 4.8 Atmospheric fronts are narrow regions of large horizontal temperature gradient. A simple model takes the front to be a sloping surface across which the temperature and along-front wind are discontinuous (with warm air overlying cold), but the pressure and cross-front wind are continuous. Take the y -axis along the front and the x -axis pointing towards the cold air and apply the hydrostatic and geostrophic wind relationships to the region AB in Figure 4.17. Hence show that the slope α at any level is related to the temperatures T_1 and T_2 and along-front winds v_1 and v_2 at that level by

$$(T_2 - T_1)g \tan \alpha = f(v_1 T_2 - v_2 T_1).$$

If $T_2 - T_1 = 3$ K and $v_1 - v_2 = 10 \text{ m s}^{-1}$, estimate α at a latitude of 50° N .

Problem 4.9 Starting with the geostrophic and hydrostatic equations in pressure coordinates, derive the thermal windshear relationship in the form

$$\frac{\partial}{\partial p}(u, v) = \frac{R_a}{f p} \left(\frac{\partial T}{\partial y}, -\frac{\partial T}{\partial x} \right),$$

making clear which variables are held constant in each partial derivative. How does this equation differ from the z -coordinate version?

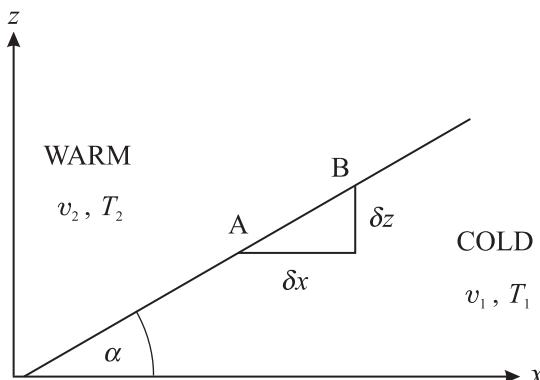


Fig. 4.17

A simple model of a front.

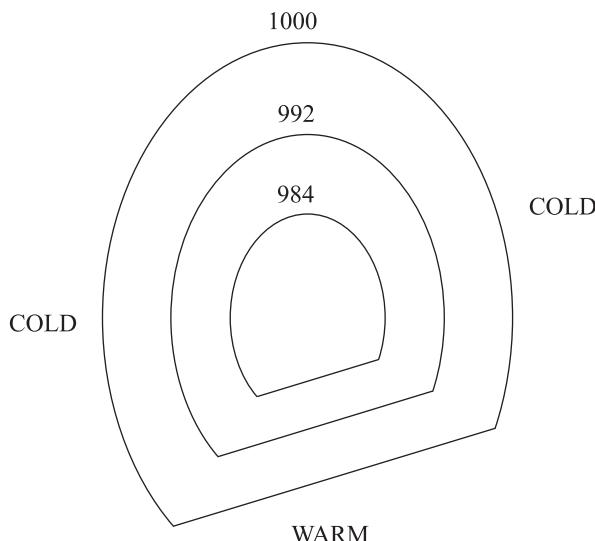


Fig. 4.18

A schematic weather map of a developing cyclone.

Introduce the variable $Z = \ln(p_0/p)$, where p_0 is a constant reference pressure, and derive the thermal wind relationship using Z as a vertical coordinate.

The temperatures in the following table are derived from satellite measurements for latitudes 50° N and 40° N and longitude 0° E. Given that the eastward wind at $(45^\circ$ N, 0° E) and a pressure of 10 hPa is 25 m s^{-1} , estimate the eastward wind at $(45^\circ$ N, 0° E) at a pressure of 1 hPa.

	Pressure (hPa)	10	1
Temperature at 50° N (K)	217	252	
Temperature at 40° N (K)	224	261	

(Use the trapezoidal rule $\int_a^b F(Z) dZ \approx \frac{1}{2}[F(a) + F(b)](b - a)$.)

Problem 4.10 Figure 4.18 is a schematic ‘weather map’ of a developing cyclone (in the Northern Hemisphere) showing contours (isobars) of surface pressure in hPa and the regions of cold and warm air at the surface (the latter is called the ‘warm sector’). Giving physical justification, mark the directions of the geostrophic winds by arrows and the positions of the cold (▲▲▲) and warm (■■■■) fronts. (Note: warm air follows cold in a warm front and vice versa in a cold front.) Check that the changes in wind direction at the fronts agree with the results of Problem 4.8.

Problem 4.11 Figure 4.19 is part of a North Atlantic weather chart, for the sector 10° W– 35° W and 35° N– 55° N, showing surface pressure (dashed contours, in hPa) and the height of the 500 hPa surface (solid contours, in decametres; 1 decametre = 10 m). Show that, near 1000 hPa, the pressure drops by about 8 hPa for every 60 m of vertical ascent (take the density of air at the surface to be about 1.3 kg m^{-3}). Hence show that the surface

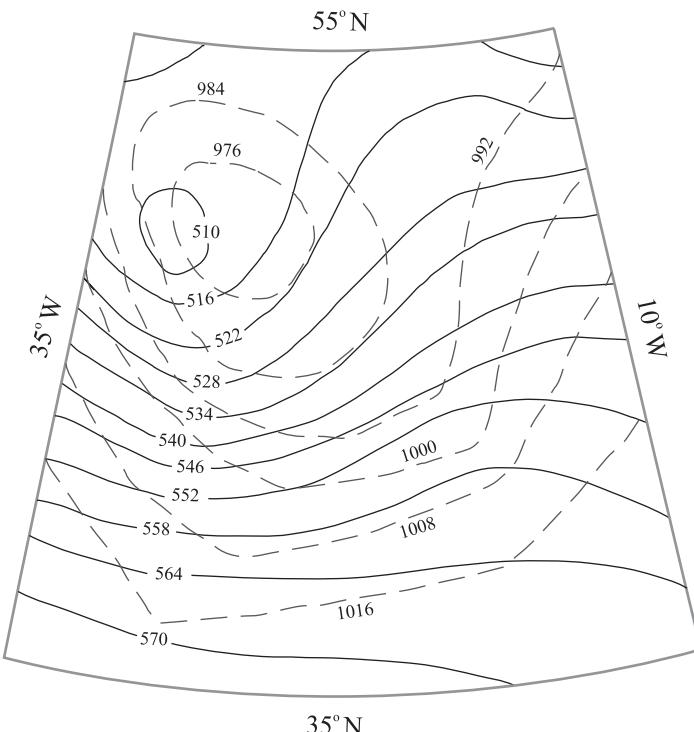


Fig. 4.19

Part of a North Atlantic weather chart.

isobars can be roughly re-interpreted as isopleths of the 1000 hPa height, and re-label them accordingly on the chart (in decametres). Draw in contours of the thickness of the layer between 1000 hPa and 500 hPa, and shade the region where the mean temperature of this layer is largest. What is the name given to this region? What is its mean temperature?

Shade in the region of strongest geostrophic winds at 500 hPa and estimate the maximum geostrophic wind speed there.

In this chapter we build on the foundations laid in [Chapter 4](#) by considering some more advanced atmospheric fluid-dynamical concepts and studying simple models of observed dynamical phenomena in the atmosphere. In [Section 5.1](#) we introduce the important notions of **vorticity** and **potential vorticity**. In [Sections 5.2](#) and [5.3](#) we make some further simplifications to the basic equations of motion, which allow us to set up simple models of two types of atmospheric **wave**: the small-scale gravity wave (in [Section 5.4](#)) and the large-scale Rossby wave (in [Section 5.5](#)). As mentioned in [Chapter 1](#), waves are very important atmospheric phenomena, not only because they are a common feature of observations, but also because they allow one part of the atmosphere to ‘communicate’ with other, perhaps distant, parts of the atmosphere. In [Section 5.6](#) we look at atmospheric **boundary layers**, regions near the Earth’s surface where frictional effects become important, and in [Section 5.7](#) we briefly touch on the important topic of atmospheric **instability**.

5.1 Vorticity and potential vorticity

The vorticity $\boldsymbol{\omega}$ is defined as the curl of the velocity vector $\mathbf{u} = (u, v, w)$:

$$\boldsymbol{\omega} = \nabla \times \mathbf{u} = \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}, \quad \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}, \quad \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right).$$

In the special case of two-dimensional flow parallel to the x, y plane and independent of z , so that $\mathbf{u} = (u(x, y), v(x, y), 0)$, the vorticity has only a z component:

$$\boldsymbol{\omega} = (0, 0, \xi), \quad \text{where} \quad \xi = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}. \quad (5.1)$$

The vorticity is a measure of the local (not global) rotation or ‘spin’ of the flow.¹ This fact is best illustrated by some simple examples:

- (a) Consider two-dimensional circular flow, using plane polar coordinates r and φ , as in [Section 4.8.2](#), so that $\mathbf{u} = V(r)\mathbf{i}_\varphi$, where \mathbf{i}_φ is the unit vector in the azimuthal direction. It can be shown (see [Problem 5](#)/[Problem 5.1](#)) that the vorticity is in the z direction, as in [equation \(5.1\)](#), with

$$\xi = \frac{dV}{dr} + \frac{V}{r}. \quad (5.2)$$

¹ The word ‘spin’ is used in the sense of classical, not quantum, mechanics here.

Consider two special cases.

- ‘Solid-body rotation’, in which the fluid rotates with the same angular velocity Ω at all points, so that $V = \Omega r$. In this case equation (5.2) implies that $\xi = 2\Omega = \text{constant}$.
 - The ‘point vortex’, in which $V \propto r^{-1}$; in this case $\xi = 0$. (The origin $r = 0$ is a singular point and must be excluded.) This is an example of a flow that we can clearly regard as rotating in a global sense, but that has zero vorticity. Thus vorticity is not a signature of global rotation.
- (b) Now consider two-dimensional rectilinear shear flow in Cartesian coordinates: $\mathbf{u} = (u(y), 0, 0)$. In this case $\xi = -du/dy$ and is generally non-zero. Such a flow may be regarded as ‘non-rotating’ in a global sense, but has non-zero vorticity. Again vorticity is not directly linked with global rotation.

Vorticity as a local quantity may usefully be pictured in the following way. Imagine a tiny hypothetical ‘vorticity meter’, consisting of four perpendicular vanes (see Figure 5.1), which is placed in a two-dimensional flow with its axis in the z direction. The vanes tend to be carried with the local flow; one vane is marked with a black dot, as shown. Close examination shows that the angular velocity of the meter is equal to half the local value of the vorticity.

When it is placed in the solid-body rotation, the meter revolves in a circular orbit, as shown in Figure 5.2(a). Moreover, the azimuthal flow near the outer vane is faster than that near the inner vane, by just such an amount as to give the vane a single rotation per orbit. The meter spins with angular velocity Ω , equal to half the vorticity ξ .

When it is placed in the point vortex, however, the meter remains aligned with its original orientation as it orbits; see Figure 5.2(b). In this case the flow near the outer vane is less than that near the inner vane, by just such an amount as to keep the meter aligned. The meter has zero spin, which is consistent with the zero vorticity of the flow.

In the case of the rectilinear shear flow (Figure 5.2(c)), with $u = By$, where B is a positive constant, the flow near the upper vane in the diagram is faster than that near the lower vane. The meter spins in a clockwise sense, with negative angular velocity, which is consistent with the vorticity ξ being $-B$ here.

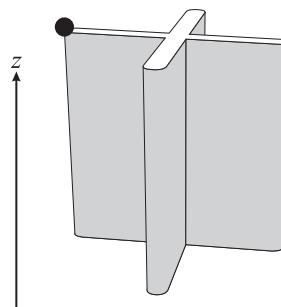


Fig. 5.1

A hypothetical ‘vorticity meter’. Adapted from Acheson (1990).

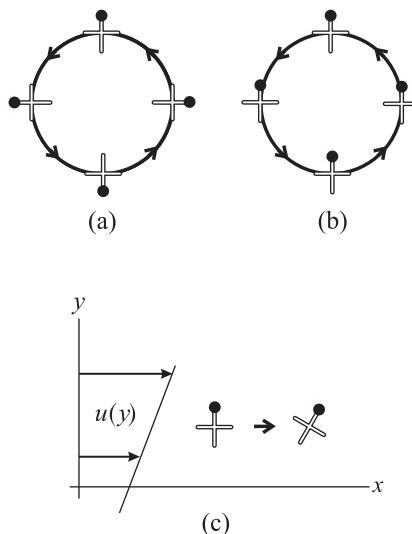


Fig. 5.2 The behaviour of a hypothetical vorticity meter in (a) solid-body rotation, (b) a point vortex and (c) rectilinear shear flow. Adapted from [Acheson \(1990\)](#).

The concept of vorticity may be extended to rotating frames. Putting $\mathbf{A} = \mathbf{r}$ in [equation \(4.18\)](#), we obtain the relationship between the velocities in an inertial frame and a rotating frame:

$$\mathbf{u}_I = \mathbf{u}_R + \boldsymbol{\Omega} \times \mathbf{r},$$

where \mathbf{u}_I is the velocity in the inertial frame and \mathbf{u}_R is the velocity in the rotating frame. Taking the curl and using the vector identity

$$\nabla \times (\mathbf{a} \times \mathbf{b}) = (\mathbf{b} \cdot \nabla)\mathbf{a} - (\mathbf{a} \cdot \nabla)\mathbf{b} + \mathbf{a}(\nabla \cdot \mathbf{b}) - \mathbf{b}(\nabla \cdot \mathbf{a}), \quad (5.3)$$

where \mathbf{a} and \mathbf{b} are vector functions of position, we obtain

$$\boldsymbol{\omega}_I = \boldsymbol{\omega}_R + 2\boldsymbol{\Omega}. \quad (5.4)$$

Thus the vorticity vector in the inertial frame (the **absolute vorticity**) is that in the rotating frame (the **relative vorticity**) plus twice the rotation vector. This is consistent with the two-dimensional solid-body rotation result above, where we can regard the flow as being at rest in a frame rotating with angular speed $\boldsymbol{\Omega}$.

The concept of **potential vorticity** is more complex and will only be touched on here. The potential vorticity (sometimes called the **Rossby–Ertel potential vorticity**) is defined as

$$P = \frac{\boldsymbol{\omega}_I \cdot \nabla \theta}{\rho}, \quad (5.5)$$

where θ is the potential temperature and ρ the fluid density. Given the Navier–Stokes equation [\(4.20\)](#), the mass-continuity [equation \(4.8\)](#) and the thermodynamic [equation \(4.37\)](#),

it can be shown that the potential vorticity satisfies an equation of the form

$$\frac{DP}{Dt} = \text{terms involving friction and diabatic heating.}$$

A corollary of this result is that, in the absence of friction and diabatic heating, the right-hand side vanishes and the potential vorticity is **materially conserved**: in other words, each fluid blob retains a fixed value of potential vorticity as it moves around and the potential vorticity acts as a **tracer** of fluid motion. Together with other properties, this makes the potential vorticity a very useful fluid-flow quantity, which has been used in numerous studies as a diagnostic of atmospheric motion.

5.2 The Boussinesq approximation

In Chapter 4 we derived the following set of equations for atmospheric flow, valid on a sphere, f -plane or β -plane: the momentum equations

$$\frac{Du}{Dt} - fv + \frac{1}{\rho} \frac{\partial p}{\partial x} = F^{(x)}, \quad (5.6a)$$

$$\frac{Dv}{Dt} + fu + \frac{1}{\rho} \frac{\partial p}{\partial y} = F^{(y)}, \quad (5.6b)$$

$$\frac{Dw}{Dt} + \frac{1}{\rho} \frac{\partial p}{\partial z} + g = 0, \quad (5.6c)$$

the mass-continuity equation

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{u} = 0, \quad (5.6d)$$

the ideal gas law

$$p = RT\rho \quad (5.6e)$$

and the thermodynamic energy equation

$$c_p \frac{DT}{Dt} - \frac{1}{\rho} \frac{Dp}{Dt} = Q. \quad (5.6f)$$

In equations (5.6a) and (5.6b) $f = 2\Omega \sin \phi$ if spherical coordinates are used, $f = f_0$ on an f -plane and $f = f_0 + \beta y$ on a β -plane.

These equations are still quite complicated, so we now introduce the **Boussinesq approximation**, which further simplifies the mathematics while retaining much of the important physics. It is motivated by the fact that compressibility effects in the atmosphere may be neglected for many purposes, except when considering layers whose depths are greater than the density scale height, over which the density falls by a factor of e . In particular, sound waves may be ignored for meteorological purposes. Taking the atmosphere to be

incompressible (see the end of Section 4.3) implies that equation (5.6d) decouples into two equations,

$$\nabla \cdot \mathbf{u} = 0, \quad (5.7)$$

$$\frac{D\rho}{Dt} = 0. \quad (5.8)$$

Equation (5.8) states that, for incompressible flow, the density is constant on following a moving fluid blob. However, this does not imply that the density is uniform everywhere; we must still allow for vertical density stratification. It is convenient to separate the density into a ‘background’ part $\bar{\rho}$ that depends only on height z and a deviation ρ' :

$$\rho(x, y, z, t) = \bar{\rho}(z) + \rho'(x, y, z, t) \quad (5.9)$$

with a similar separation for the pressure:

$$p(x, y, z, t) = \bar{p}(z) + p'(x, y, z, t).$$

We now approximate the vertical momentum equation (5.6c) by hydrostatic balance, as in equation (4.25):

$$\frac{\partial p}{\partial z} = -g\rho. \quad (5.10)$$

If we assume that the background state itself satisfies hydrostatic balance,

$$\frac{d\bar{p}}{dz} = -g\bar{\rho},$$

then equation (5.10) implies that the deviation also satisfies hydrostatic balance:

$$\frac{\partial p'}{\partial z} = -g\rho'. \quad (5.11)$$

Substitution of equation (5.9) into the density equation (5.8) leads to

$$\frac{D\rho'}{Dt} + w \frac{d\bar{\rho}}{dz} = 0. \quad (5.12)$$

This states that the density deviation ρ' of a blob changes as the blob moves up or down in the background density gradient $d\bar{\rho}/dz$. By analogy with equation (2.31) we introduce the quantity

$$N_B^2 = -\frac{g}{\rho_0} \frac{d\bar{\rho}}{dz}, \quad (5.13)$$

where $\rho_0 = \bar{\rho}(0)$, say, a reference value of the background density. $N_B(z)$ is the buoyancy frequency for the stratified, incompressible fluid. Then equation (5.12) can be written

$$g \frac{D\rho'}{Dt} - \rho_0 N_B^2 w = 0. \quad (5.14)$$

The **Boussinesq approximation** is now implemented by ignoring density variations except where they are coupled with gravity; that is, we replace ρ by the constant value ρ_0 in the horizontal momentum equations (5.6a) and (5.6b) but retain the full density variation

(and in particular the deviation ρ') in equations (5.11) and (5.14). We therefore obtain the following **Boussinesq equations**:²

$$\frac{Du}{Dt} - fv + \frac{1}{\rho_0} \frac{\partial p}{\partial x} = F^{(x)}, \quad (5.15a)$$

$$\frac{Dv}{Dt} + fu + \frac{1}{\rho_0} \frac{\partial p}{\partial y} = F^{(y)}, \quad (5.15b)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \quad (5.15c)$$

$$\frac{D}{Dt} \left(-\frac{g\rho'}{\rho_0} \right) + N_B^2 w = 0, \quad (5.15d)$$

$$\frac{\partial p'}{\partial z} = -g\rho'. \quad (5.15e)$$

It should be noted that, under the Boussinesq approximation, the thermodynamics becomes decoupled from the dynamics. If equations (5.15) are solved, however, we can obtain the temperature using the ideal gas law

$$T = \frac{\bar{p} + p'}{R_a(\bar{\rho} + \rho')}.$$

A ‘mass source’ term is sometimes added to the right-hand side of equation (5.15d), by analogy with the diabatic heating Q that appears in the thermodynamic energy equation.

5.2.1 Linearised equations and energetics

Because of the presence of the $(\mathbf{u} \cdot \nabla)$ term in the material derivative D/Dt , equations (5.15) are non-linear and therefore difficult to solve analytically. Non-linear equations of this type are routinely solved on computers, for example in weather forecasting; however, numerical solutions are often difficult to interpret physically. We therefore now make the further approximation of **linearising** these equations, so that solutions can be found and analysed.

We assume that velocities and density deviations are ‘small’, in the sense that terms that are quadratic in these quantities and their derivatives can be neglected. For example, equation (5.15d) can be expanded as

$$-\frac{g}{\rho_0} \left(\underline{\frac{\partial \rho'}{\partial t}} + \underline{u \frac{\partial \rho'}{\partial x}} + \underline{v \frac{\partial \rho'}{\partial y}} + \underline{w \frac{\partial \rho'}{\partial z}} \right) + N_B^2 w = 0,$$

where the quadratic terms are underlined. If these are dropped we get

$$-\frac{g}{\rho_0} \left(\frac{\partial \rho'}{\partial t} \right) + N_B^2 w = 0,$$

² These equations are very similar in form to the full equations in pressure coordinates: see Section 4.9.

which is *linear* in ρ' and w . The full set of Boussinesq equations, linearised in this way and with friction neglected, is

$$u_t - fv + \frac{1}{\rho_0} p'_x = 0, \quad (5.16a)$$

$$v_t + fu + \frac{1}{\rho_0} p'_y = 0, \quad (5.16b)$$

$$u_x + v_y + w_z = 0, \quad (5.16c)$$

$$-\frac{g}{\rho_0} \rho'_t + N_B^2 w = 0, \quad (5.16d)$$

$$p'_z + g\rho' = 0. \quad (5.16e)$$

(The incompressibility condition (5.16c) and hydrostatic equation (5.16e) were already linear and needed no further approximation here.) Note that we have introduced the useful shorthand notation of denoting partial derivatives with respect to x, y, z and t by subscripts.

An important result can be obtained by multiplying equation (5.16a) by $\rho_0 u$, equation (5.16b) by $\rho_0 v$ and equation (5.16d) by $-g\rho'/N_B^2$ and then adding the results. A short calculation gives the **energy equation**

$$\frac{\partial}{\partial t} \frac{1}{2} \rho_0 \left[u^2 + v^2 + \left(\frac{g\rho'}{\rho_0 N_B} \right)^2 \right] + \nabla \cdot (\mathbf{u} p') = 0. \quad (5.17)$$

In this equation, $\rho_0(u^2 + v^2)/2$ is clearly the kinetic energy per unit volume of the horizontal motion, while the term involving $(\rho')^2$ can be identified as the available potential energy (with respect to a reference state at rest, with density $\bar{\rho}$ and pressure \bar{p} : see Section 2.6). The term $\mathbf{u} p'$ can be interpreted as an energy flux (analogous to the Poynting vector in electromagnetism). Overall, the equation states that the energy (kinetic plus available potential) within a volume increases if there is an energy flux into the volume and decreases if there is an energy flux out of the volume. Similar, but more complicated, versions of equation (5.17) apply for the nonlinear Boussinesq equations (5.15) and the nonlinear compressible equations (5.6).

5.3 Quasi-geostrophic motion

In Section 4.8 we derived the geostrophic approximations to equations (5.6a) and (5.6b). Under the Boussinesq approximation, neglecting variations of f , geostrophic balance can be expressed by

$$u \approx u_g = -\frac{\partial \psi}{\partial y}, \quad v \approx v_g = \frac{\partial \psi}{\partial x}, \quad (5.18)$$

where ψ is the **geostrophic streamfunction**, defined by

$$\psi \equiv \frac{p'}{f_0 \rho_0}. \quad (5.19)$$

(See equations (4.26a), (4.26b) and note that we can use p' rather than p in equation (5.19) since the horizontal derivatives of $\bar{p}(z)$ vanish.) The velocity vector $(u_g, v_g, 0)$ is called the **geostrophic flow**. From equations (5.18)

$$\frac{\partial u_g}{\partial x} + \frac{\partial v_g}{\partial y} = 0,$$

so by comparison with the incompressibility condition (5.15c) we can see that any vertical velocity w_g associated with the geostrophic flow must be independent of z . It is therefore zero everywhere if it vanishes at one level, say, at the ground $z = 0$.

Note that, using the geostrophic streamfunction ψ , hydrostatic balance (5.15e) can be written as

$$\rho' = -\frac{f_0 \rho_0}{g} \frac{\partial \psi}{\partial z}. \quad (5.20)$$

We now seek a better approximation to the nonlinear Boussinesq equations (5.15) than that given by pure geostrophic balance. One way to do this is to introduce the **ageostrophic velocity**, the difference between the true velocity and the geostrophic flow, with components

$$u_a \equiv u - u_g, \quad v_a \equiv v - v_g, \quad w_a \equiv w.$$

It can then be shown that the next approximation beyond geostrophic balance is given, on a β -plane and neglecting friction, by the **quasi-geostrophic equations**

$$D_g u_g - f_0 v_a - \beta y v_g = 0, \quad (5.21a)$$

$$D_g v_g + f_0 u_a + \beta y u_g = 0, \quad (5.21b)$$

$$\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y} + \frac{\partial w_a}{\partial z} = 0, \quad (5.21c)$$

$$D_g \left(-\frac{g \rho'}{\rho_0} \right) + N_B^2 w_a = 0. \quad (5.21d)$$

Here

$$D_g \equiv \frac{\partial}{\partial t} + u_g \frac{\partial}{\partial x} + v_g \frac{\partial}{\partial y}$$

is the time derivative following the geostrophic flow. The quasi-geostrophic equations hold in general for large-scale, low-frequency motions, except in low latitudes; in particular they require that the Rossby number (see equation (4.27)) should be small:

$$\text{Ro} \equiv \frac{U}{fL} \ll 1, \quad (5.22)$$

and that time scales should be large compared with $1/f_0$ (a few hours, except in low latitudes). They provide a useful model for investigating many types of large-scale motion that are observed in the atmosphere.

Equations (5.21) can conveniently be combined as follows. First take $\partial/\partial x$ (5.21b) – $\partial/\partial y$ (5.21a) and use the mass-continuity equation (5.21c), to obtain (after careful

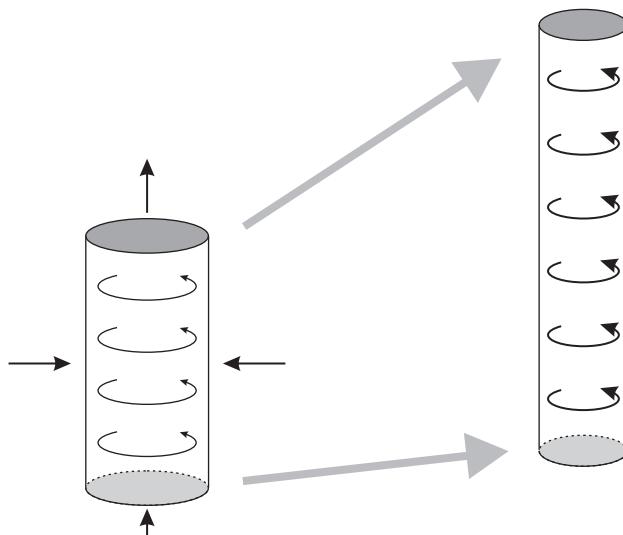


Fig. 5.3

Illustrating the ‘stretching’ mechanism of absolute vorticity generation. The cylindrical blob of fluid on the left experiences stretching in the vertical owing to the differential vertical velocities (indicated by the vertical arrows), and shrinking in the horizontal associated with convergent horizontal velocities (indicated by the horizontal arrows). Its vertical absolute vorticity is indicated by the curved arrows. It is deformed by the velocity field into the shape shown on the right, and at the same time its vertical absolute vorticity increases.

manipulation) the **vorticity equation**

$$D_g \zeta = f_0 \frac{\partial w_a}{\partial z}, \quad (5.23)$$

where

$$\zeta \equiv f_0 + \beta y - \frac{\partial u_g}{\partial y} + \frac{\partial v_g}{\partial x} = f_0 + \beta y + \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2}$$

is the z component of the **absolute vorticity** associated with the geostrophic flow; cf. equation (5.4).

The term on the right-hand side of equation (5.23) is called the **stretching term**, since it can generate vorticity by differential vertical motion. Consider a cylindrical blob of air, moving with the geostrophic flow: see Figure 5.3. If the blob enters a region where the vertical velocity w_a increases with height, it is stretched vertically. Since its mass must be conserved it shrinks in the horizontal direction,³ consistent with there being a horizontal convergence of velocity; see equation (5.21c). Equation (5.23) states that the z component of the absolute vorticity ζ of the blob must increase (in the Northern Hemisphere, where $f_0 > 0$) by this stretching mechanism. Conversely a cylindrical blob that is squashed in the vertical direction will suffer a decrease in ζ .

³ Remember that, under the Boussinesq approximation, the fluid is incompressible.

Now note that, using equations (5.21d) and (5.20), the vertical velocity can be expressed as

$$w_a = D_g \left(\frac{g\rho'}{\rho_0 N_B^2} \right) = -D_g \left(\frac{f_0}{N_B^2} \frac{\partial \psi}{\partial z} \right); \quad (5.24)$$

substitution of this expression into the vorticity equation (5.23) and further careful manipulation lead finally to the **quasi-geostrophic potential vorticity equation**

$$D_g q = 0, \quad (5.25)$$

where

$$q \equiv \zeta + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N_B^2} \frac{\partial \psi}{\partial z} \right) = f_0 + \beta y + \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N_B^2} \frac{\partial \psi}{\partial z} \right) \quad (5.26)$$

is the **quasi-geostrophic potential vorticity** (QGPV). When friction and diabatic heating (or mass sources) are neglected, as here, the QGPV is constant following the geostrophic flow. We can ‘forecast’ the future behaviour of the QGPV by integrating equation (5.25) forwards in time if the current values of the QGPV and the geostrophic flow are known at all points. Moreover, the QGPV carries information about the geostrophic flow and the density, since the elliptic operator (analogous to ∇^2) on the right-hand side of equation (5.26) can be inverted, given suitable boundary conditions, to give ψ and hence p' , u_g , v_g and ρ' , using equations (5.19), (5.18) and (5.20).

The QGPV is somewhat analogous to the Rossby–Ertel potential vorticity (5.5), which is, however, constant following the *total* flow \mathbf{u} in the absence of friction and diabatic heating.

5.4 Gravity waves

It was noted in Chapter 1 that wave-like motions are frequently observed in the atmosphere. One such wave is the **gravity wave**: examples of gravity waves include the lee waves that are often manifested as parallel bands of cloud downstream of mountain ranges; see Figure 5.4. Another example, derived from ground-based radar measurements and demonstrating ‘waviness’ (though not precisely sinusoidal behaviour) in time, is given in Figure 1.7.

The linear Boussinesq equations (5.16) are a suitable set from which to develop models of gravity waves. As in modelling other wave motions in physics (e.g. electromagnetic waves as solutions of Maxwell’s equations), it is simplest to look for **plane-wave solutions** of equations (5.16). First, however, we restrict our attention to motions of comparatively small horizontal scale (for example, $\lesssim 100$ km) so that the rotation of the Earth has a negligible effect. This allows us to simplify the equations by neglecting the Coriolis terms (i.e. by putting $f = 0$) in equations (5.16a) and (5.16b). We also assume that N_B is independent of z : this is a fairly reasonable assumption both for the troposphere and for the stratosphere.

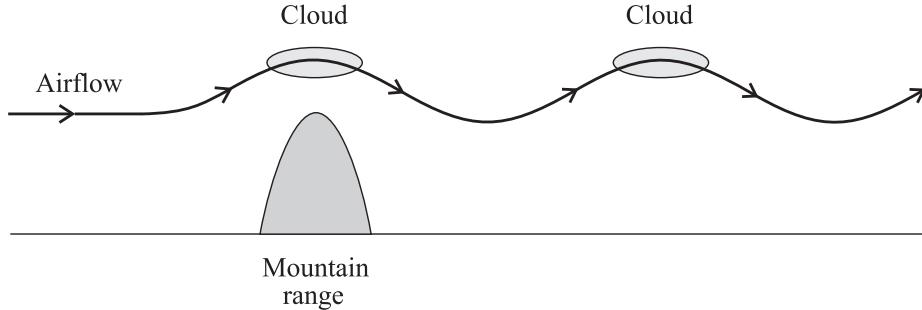


Fig. 5.4

A schematic cross-section through a lee wave (or mountain wave), an internal gravity wave that may form when a stratified airflow blows over a mountain range. The wavy arrow indicates one particular streamline of the flow. The wave motion appears over and downstream of the mountain and may be made visible by clouds that form as water vapour in the moving air parcels condenses in the rising parts of the flow. The clouds disperse as the air parcels descend and the water evaporates again. Note that the wave pattern and clouds are stationary with respect to the mountain, but that the air-flow blows through them.

We seek plane waves, propagating in the x, z plane and independent of y , of the form

$$\{u, v, w, p', \rho'\} = \text{Re} \left\{ \hat{u}, \hat{v}, \hat{w}, \hat{p}, \hat{\rho} \right\} \exp[i(kx + mz - \omega t)], \quad (5.27)$$

where \hat{u} , etc. are complex amplitudes. Substitution of expressions (5.27) into the linear partial differential equations (5.16) then yields the algebraic equations

$$-i\omega\hat{u} + ik\hat{p}/\rho_0 = 0, \quad (5.28a)$$

$$-i\omega\hat{v} = 0, \quad (5.28b)$$

$$ik\hat{u} + im\hat{w} = 0, \quad (5.28c)$$

$$i\omega g\hat{\rho}/\rho_0 + N_B^2 \hat{w} = 0, \quad (5.28d)$$

$$im\hat{p} + g\hat{\rho} = 0. \quad (5.28e)$$

It is straightforward to eliminate, say, \hat{u} , \hat{v} , \hat{w} and $\hat{\rho}$ in favour of \hat{p} ; arbitrarily choosing \hat{p} to be real, we then get

$$p' = \hat{p} \cos(kx + mz - \omega t), \quad (5.29a)$$

$$u = \frac{k\hat{p}}{\rho_0 \omega} \cos(kx + mz - \omega t), \quad (5.29b)$$

$$v = 0, \quad (5.29c)$$

$$w = -\frac{k^2 \hat{p}}{\rho_0 \omega m} \cos(kx + mz - \omega t), \quad (5.29d)$$

$$\rho' = \frac{m\hat{p}}{g} \sin(kx + mz - \omega t). \quad (5.29e)$$

These are called the **polarisation relations** for the waves. Note that u , w and p' are in phase (or 180° out of phase) with each other and 90° out of phase with ρ' , and that $v = 0$ for these y -independent waves. The condition for the algebraic equations (5.28) to possess a non-trivial solution gives the **dispersion relation**, relating the angular frequency ω to the wave-vector components k and m , for **internal gravity waves**:

$$\omega^2 = \frac{N_B^2 k^2}{m^2}. \quad (5.30)$$

On taking the square root we obtain the two possible solutions

$$\omega = \pm \frac{N_B k}{m}. \quad (5.31)$$

To understand the physical difference between these two solutions, we must introduce the vector group velocity

$$\mathbf{c}_g = (c_g^{(x)}, 0, c_g^{(z)}) = \left(\frac{\partial \omega}{\partial k}, 0, \frac{\partial \omega}{\partial m} \right); \quad (5.32)$$

in particular the vertical component of the group velocity, $c_g^{(z)}$, is

$$c_g^{(z)} = \frac{\partial}{\partial m} \left(\pm \frac{N_B k}{m} \right) = \mp \frac{N_B k}{m^2}. \quad (5.33)$$

We now adopt the convention that k is positive.⁴ For an atmospheric internal gravity wave generated near the ground and propagating information upwards, we must have $c_g^{(z)} > 0$, so the lower sign must be chosen both in equation (5.33) and in equation (5.31), which then becomes

$$\omega = -\frac{N_B k}{m}.$$

If also $\omega > 0$, we have $m < 0$ and the phase relations between the velocity, density and pressure disturbances due to the waves may be summarised as in Figure 5.5.

Several features should be noted here.

- For this choice of signs, the phase surfaces $kx + mz - \omega t = \text{constant}$ move obliquely downwards in time, in the direction of the wave-vector $\mathbf{k} = (k, 0, m)$. However, the propagation of information, represented by the group velocity vector \mathbf{c}_g , is obliquely upwards.
- The velocity vector $(u, 0, w)$ is parallel to the slanting phase surfaces: fluid blobs (or parcels) oscillate up and down these surfaces.
- The precise phase relations between the velocity, density and pressure disturbances may be verified from the polarisation relations (5.29).

The hydrostatic assumption is equivalent to neglecting a term $\rho_0 \partial w / \partial t$ compared with $g\rho'$, say, in the linear equation (5.16e). From the polarisation relations (5.29d) and (5.29e)

⁴ The following analysis may be repeated with the convention $k < 0$; certain changes of sign must be made in the analysis, but the physical content of the resulting equations is unchanged.

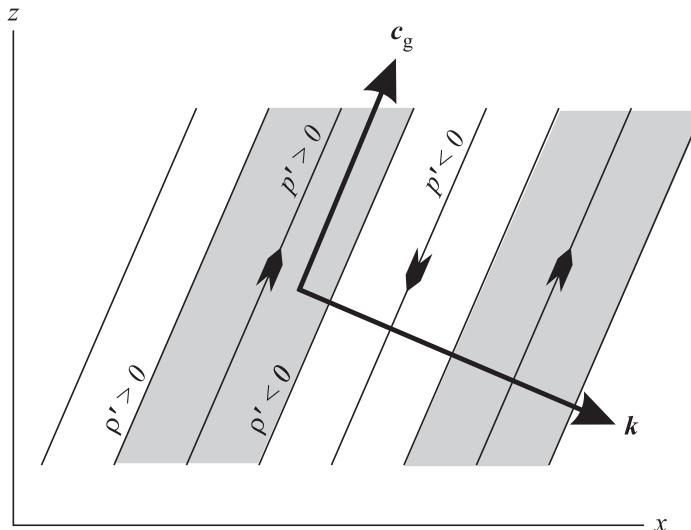


Fig. 5.5

A vertical cross-section through a plane internal gravity wave with $k > 0$, $\omega > 0$ and $m < 0$ so that $c_g^{(z)} > 0$, as in the text. The thin sloping lines represent surfaces of constant phase, separated perpendicularly from one another by a quarter-wavelength. The phase surfaces on which the density and pressure disturbances take their greatest positive and negative values are marked. The phase surfaces in which the velocity vectors (which are themselves parallel to the phase surfaces) have their greatest upward and downward components are marked by long black arrowheads in the appropriate directions. The regions of upward motion are shaded. The thick arrow sloping obliquely upwards indicates the direction of the group velocity vector c_g and the thick arrow sloping obliquely downwards indicates the direction of the wave-vector $k = (k, 0, m)$. The phase surfaces move perpendicular to themselves, in the direction of k , as time progresses.

it can be seen that this is valid if $k^2 \ll m^2$, i.e. if vertical wavelengths are much less than horizontal wavelengths. Equation (5.30) then shows that $\omega^2 \ll N_B^2$ under this condition, so the hydrostatic internal gravity waves we have been considering must have angular frequencies much less than the buoyancy frequency N_B . Even for non-hydrostatic waves it turns out that $\omega^2 \leq N_B^2$; typical values of the minimum period $2\pi/N_B$ are 8 min for the troposphere and 5 min for the stratosphere.

A useful connection can be made here with the analysis of Section 2.5, in which the buoyancy frequency N in a compressible, stably stratified fluid was shown to be the frequency at which a parcel (or blob) of fluid oscillates vertically up and down. We have just seen that, in an internal gravity wave, fluid blobs do not move purely vertically, but are constrained to move along the sloping phase surfaces. These surfaces make an angle α with the horizontal, where $\sin^2 \alpha = k^2/(k^2 + m^2)$; see Figure 5.6. Consider a blob of unit mass: under the Boussinesq approximation its volume V remains constant and its density is approximately ρ_0 , so that $\rho_0 V \approx 1$. If this blob moves a vertical distance δz upwards from its equilibrium position and hence a distance $\delta z/\sin \alpha$ along the slope, then its density is $\rho_0 N_B^2 \delta z/g$ greater than that of its surroundings, by equation (5.13). It therefore experiences

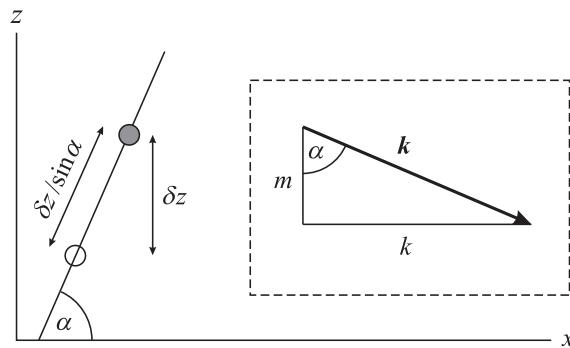


Fig. 5.6

An illustration of the motion of a fluid blob in a plane internal gravity wave. The sloping line on the left of the diagram indicates a sloping phase surface. The open circle indicates the equilibrium position of the blob and the solid circle its displaced position. The inset triangle on the right of the diagram shows the wave-vector $\mathbf{k} = (k, 0, m)$, which is perpendicular to the phase surface.

a downward buoyancy force $N_B^2 \delta z$, which has a component $-N_B^2 \sin \alpha \delta z$ up the sloping phase surface. Its acceleration up the slope is

$$\frac{d^2}{dt^2} \left(\frac{\delta z}{\sin \alpha} \right),$$

so, by Newton's Second Law,

$$\frac{d^2}{dt^2} \left(\frac{\delta z}{\sin \alpha} \right) + N_B^2 \sin \alpha \delta z = 0$$

(cf. equation (2.30)). This implies an oscillation whose angular frequency ω is given by

$$\omega^2 = N_B^2 \sin^2 \alpha = \frac{N_B^2 k^2}{k^2 + m^2} \approx \frac{N_B^2 k^2}{m^2}$$

(the approximation holding for hydrostatic waves), in agreement with the dispersion relation (5.30).

In this section we have concentrated on internal gravity waves, the class of gravity waves whose horizontal scales are so small that the Earth's rotation can be neglected. A similar study can be carried out for waves of somewhat larger scale (horizontal wavelengths of hundreds of kilometres) and somewhat lower frequency (periods of several hours), which do feel the Earth's rotation; these are called **inertia-gravity waves**. (The waves in Figure 1.7 are of this type.) An example is given in Problem 5.4.

5.5 Rossby waves

A further class of atmospheric wave is observed on horizontal scales of thousands of kilometres and with periods of several days. These are known as **Rossby waves** or **planetary waves**: an example is shown in Figure 1.8. These are even more difficult to identify

unambiguously than gravity waves, since a purely sinusoidal Rossby-wave structure can exist only in very simple background flows. However, they are conceptually very important for our understanding of many large-scale atmospheric phenomena.

Since atmospheric Rossby waves occur on large horizontal wavelengths and with low frequencies, they have low Rossby numbers (see equations (4.27) and (5.22)). It is therefore reasonable to try to model them using the quasi-geostrophic equations developed in Section 5.3. We consider small-amplitude disturbances to a uniform zonal background flow $(U, 0, 0)$, where U is a constant; by equation (5.18) this uniform flow corresponds to a geostrophic streamfunction $\Psi = -Uy$. For the total flow (the background plus a small disturbance) we therefore take

$$\psi = -Uy + \psi',$$

substitute into the QGPV equation (5.25) and neglect terms that are quadratic in ψ' .

The QGPV for this flow is

$$q = f_0 + \beta y + \mathcal{L}\psi', \quad (5.34)$$

where \mathcal{L} is the elliptic operator

$$\mathcal{L} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N_B^2} \frac{\partial}{\partial z} \right),$$

and the QGPV equation (5.25) linearises to

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \mathcal{L}\psi' + \beta \frac{\partial \psi'}{\partial x} = 0. \quad (5.35)$$

We take N_B to be constant, for simplicity, and look for plane-wave solutions to this equation as in Section 5.4 (but now allowing variations in y as well) by substituting

$$\psi' = \text{Re } \hat{\psi} \exp[i(kx + ly + mz - \omega t)],$$

where $\hat{\psi}$ is a complex amplitude. We obtain the dispersion relation for Rossby waves:

$$\omega = kU - \frac{\beta k}{k^2 + l^2 + f_0^2 m^2 / N_B^2}. \quad (5.36)$$

In simple modelling it is common for k , l and ω to be specified in advance, so that the vertical wavenumber m can then be found from equation (5.36):

$$m = \pm \frac{N_B}{f_0} \left(\frac{\beta}{U - (\omega/k)} - (k^2 + l^2) \right)^{1/2}. \quad (5.37)$$

Several features are immediately evident from the dispersion relation (5.36):

- The β -effect is crucial to the existence of these waves: putting $\beta = 0$ in equation (5.36) gives $\omega = kU$, which corresponds to ‘waves’ that are merely carried along with the background flow U .
- Equation (4.24) shows that $\beta > 0$. The zonal phase speed of the waves,

$$c \equiv \frac{\omega}{k} = U - \frac{\beta}{k^2 + l^2 + f_0^2 m^2 / N_B^2},$$

therefore always satisfies

$$U - c > 0; \quad (5.38)$$

that is, the wave crests and troughs (which move with the phase speed) move *westward* with respect to the background flow.

- For given real values of the horizontal wavenumbers k and l , we get vertical propagation if m is real and non-zero (and vertical evanescence if m is imaginary). Vertical propagation therefore corresponds to $m^2 > 0$ and this implies that

$$U - c = \frac{\beta}{k^2 + l^2 + f_0^2 m^2 / N_B^2} < U_c \equiv \frac{\beta}{k^2 + l^2}, \quad (5.39)$$

where U_c depends on the horizontal wavelengths of the wave. Putting equations (5.38) and (5.39) together, we therefore find that for vertical propagation we must have

$$0 < U - c < U_c.$$

In particular, for *stationary* waves, whose crests and troughs do not move with respect to the ground (which will be true for waves forced by the background flow moving over continent-scale topography or stationary heat sources) so that $c = 0$, we obtain the **Charney–Drazin criterion**

$$0 < U < U_c. \quad (5.40)$$

This states that stationary waves propagate vertically only in eastward background flows ($U > 0$) that are not too strong ($U < U_c$); moreover, since U_c increases with increasing horizontal wavelength, ‘long waves’ propagate vertically under a wider range of eastward flows than do ‘short waves’. This is consistent with observations in the Northern Hemisphere winter stratosphere, where winds are eastward and stationary Rossby waves have large horizontal scales, and with observations of the summer stratosphere, where winds are westward and stationary Rossby waves are absent.

- Rossby waves with large horizontal scales have small values of $k^2 + l^2$ and their zonal phase speed approaches

$$c_{\text{long}} = U - \frac{\beta N_B^2}{f_0^2 m^2},$$

whereas Rossby waves with short horizontal scales have large values of $k^2 + l^2$ and their zonal phase speed approaches the speed U of the background flow. This strong variation of phase speed with wavelength implies that the waves are strongly **dispersive**: an initial disturbance composed of a number of different wavelengths will tend to break up, or disperse, in time, as the various wavelength components propagate away at different phase speeds.

- We can define a group velocity vector

$$\mathbf{c}_g = (c_g^{(x)}, c_g^{(y)}, c_g^{(z)}) = \left(\frac{\partial \omega}{\partial k}, \frac{\partial \omega}{\partial l}, \frac{\partial \omega}{\partial m} \right),$$

as for gravity waves (equation (5.32)), except that here ω is taken to depend on l as well as on k and m , so c_g will have a non-zero y component.⁵ In particular, we find

$$c_g^{(z)} = \frac{\partial \omega}{\partial m} = \frac{2f_0^2 \beta km}{N_B^2 \left(k^2 + l^2 + \frac{f_0^2 m^2}{N_B^2} \right)^2};$$

if we choose $k > 0$ by convention, we see that the vertical component of the group velocity is positive and that the waves propagate information upwards, if $m > 0$: this determines the choice of sign in equation (5.37). For upward-propagating waves the phase surfaces $kx + ly + mz - \omega t = \text{constant}$ slope westward with height: this slope is observed for Rossby waves in the stratosphere.

A further possibility occurs when the waves are independent of height: this is equivalent to taking $m = 0$ above. Suppose also that there is no background flow, so that $U = 0$; then equation (5.36) reduces to

$$\omega = -\frac{\beta k}{k^2 + l^2}$$

and the phase speed $c = \omega/k$ is westward, which is consistent with equation (5.38). This result can also be obtained by a physical argument, based on vorticity, as follows.

When there is no z dependence, the disturbance part of the QGPV, $\mathcal{L}\psi'$ in equation (5.34), reduces to the disturbance vorticity,

$$\xi' = \frac{\partial^2 \psi'}{\partial x^2} + \frac{\partial^2 \psi'}{\partial y^2};$$

cf. equation (5.26). The total QGPV is then $q = f_0 + \beta y + \xi'$, of which $f_0 + \beta y$ is the background contribution due to the Earth's rotation and spherical geometry, represented by the β -effect. However, from equation (5.25) q is conserved following fluid blobs (assuming that they move essentially with the geostrophic flow). Therefore a northward-moving blob, which encounters an increasing $f_0 + \beta y$, must lose some of its disturbance vorticity ξ' , whereas a southward-moving blob must gain some ξ' .

In the special case where $l = 0$, we have $u'_g = 0$ and blobs move purely in the north-south direction. Consider a line of blobs, labelled A, B, C, etc., initially lying along a line of latitude $y = y_0$; see Figure 5.7. Suppose that these blobs are displaced into the sinusoidal pattern indicated by the solid wavy line: blob A moves southwards, so its value of ξ' increases, as indicated by the anticlockwise arrow in the figure. By the QGPV inversion process mentioned at the end of Section 5.3 this induces an anticlockwise rotation in the local velocity field, as indicated by the circular arrow; in particular, blob B is encouraged to move further south. The value of ξ' associated with blob B itself increases, inducing anticlockwise rotation near B, which tends to move C southwards and A northwards again. Applying this kind of argument to each blob, we find that, after a short

⁵ For stationary waves we must put $c = \omega = 0$ *after* differentiation. Note that these waves can still propagate information, even though their phase surfaces do not move: this is another consequence of their dispersive nature.

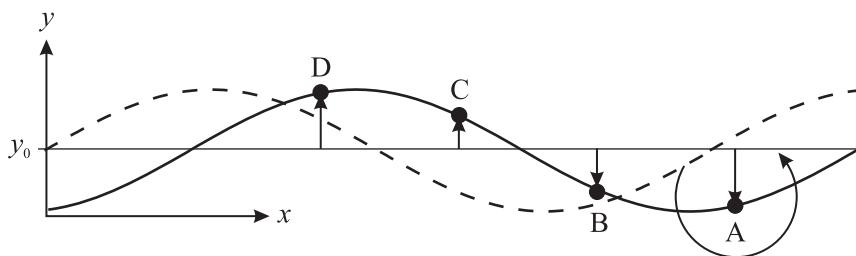


Fig. 5.7 Illustrating the Rossby-wave mechanism, in terms of conservation of potential vorticity by moving fluid blobs in the case in which the waves are independent of height. See the text for details.

time, the pattern of the blobs has moved westwards, to the position indicated by the dashed wavy line, even though each individual blob only oscillates north–south. A self-sustaining, westward-moving Rossby-wave pattern emerges, as expected from the theory given above.

5.6 Boundary layers

5.6.1 General considerations

So far in this chapter we have ignored the effects of friction in the atmosphere. This is reasonable in the models of linear internal gravity waves and Rossby waves investigated in the previous sections, since friction usually has only a small influence on these waves. However, frictional effects can sometimes be very important, especially in the lowest kilometre or so of the atmosphere. This region is sometimes called the **atmospheric boundary layer** – a terminology that echoes the use of ‘boundary layer’ in fluid dynamics to mean a thin, frictionally controlled layer near a boundary.

In the derivation of the Navier–Stokes equation (4.17) in Chapter 4 we included the effects of molecular viscosity. However, for most atmospheric purposes, the molecular viscosity is far too small to influence the dynamics directly. On the other hand, small-scale eddies can lead to a momentum transfer that in some respects resembles molecular momentum transfer. As we shall see later, this small-scale eddy transport is sometimes represented in terms of an ‘eddy viscosity’ that is similar to, but much larger than, the molecular viscosity. There are serious problems with this approach, though; for example, small-scale eddies may have organised structures, quite unlike the random nature of molecules, and may under some circumstances even imply a *negative* eddy viscosity. We shall therefore progress as far as possible without explicit use of the eddy-viscosity concept.

We consider the frictional stress τ due to small-scale processes (with scales of a few hundred metres or less) acting on the larger scales of motion in the atmosphere and assume that this stress is horizontal:

$$\tau = (\tau^{(x)}, \tau^{(y)}, 0).$$

As in equation (4.14), the vertical gradient of τ implies a horizontal force per unit mass on the larger scales, given by

$$F^{(x)} = \frac{1}{\rho_0} \frac{\partial \tau^{(x)}}{\partial z}, \quad F^{(y)} = \frac{1}{\rho_0} \frac{\partial \tau^{(y)}}{\partial z}, \quad (5.41)$$

in the Boussinesq approximation. On substituting equations (5.41) into the f -plane versions of the Boussinesq horizontal momentum equations, (5.15a) and (5.15b), and linearising, we obtain

$$\frac{\partial u}{\partial t} - f_0 v = -\frac{1}{\rho_0} \frac{\partial p}{\partial x} + \frac{1}{\rho_0} \frac{\partial \tau^{(x)}}{\partial z}, \quad (5.42a)$$

$$\frac{\partial v}{\partial t} + f_0 u = -\frac{1}{\rho_0} \frac{\partial p}{\partial y} + \frac{1}{\rho_0} \frac{\partial \tau^{(y)}}{\partial z}. \quad (5.42b)$$

We write the horizontal velocity components as

$$u = u_p + u_\tau, \quad v = v_p + v_\tau, \quad (5.43)$$

where the subscript p denotes the **pressure-driven flow**, satisfying

$$\frac{\partial u_p}{\partial t} - f_0 v_p = -\frac{1}{\rho_0} \frac{\partial p}{\partial x}, \quad (5.44a)$$

$$\frac{\partial v_p}{\partial t} + f_0 u_p = -\frac{1}{\rho_0} \frac{\partial p}{\partial y} \quad (5.44b)$$

and the subscript τ denotes the **frictional stress-driven flow**, satisfying

$$\frac{\partial u_\tau}{\partial t} - f_0 v_\tau = \frac{1}{\rho_0} \frac{\partial \tau^{(x)}}{\partial z}, \quad (5.45a)$$

$$\frac{\partial v_\tau}{\partial t} + f_0 u_\tau = \frac{1}{\rho_0} \frac{\partial \tau^{(y)}}{\partial z}. \quad (5.45b)$$

This separation can be made since equations (5.42) are linear.

We now assume that the frictional stress is important only in a boundary layer of depth D above flat ground at $z = 0$; thus $\tau^{(x)}$ and $\tau^{(y)}$ are non-zero for $0 \leq z < D$ but vanish for $z \geq D$, in the ‘free atmosphere’; see Figure 5.8. Integrating equations (5.45) through the depth of the boundary layer, we then get

$$\frac{\partial U_\tau}{\partial t} - f_0 V_\tau = -\frac{1}{\rho_0} \tau_0^{(x)}, \quad (5.46a)$$

$$\frac{\partial V_\tau}{\partial t} + f_0 U_\tau = -\frac{1}{\rho_0} \tau_0^{(y)}, \quad (5.46b)$$

where

$$U_\tau = \int_0^D u_\tau dz, \quad V_\tau = \int_0^D v_\tau dz$$

are called the **Ekman volume transports**, representing the horizontal fluxes of volume within the boundary layer, and $\tau_0^{(x)}$ and $\tau_0^{(y)}$ are the surface stresses, exerted by the ground on

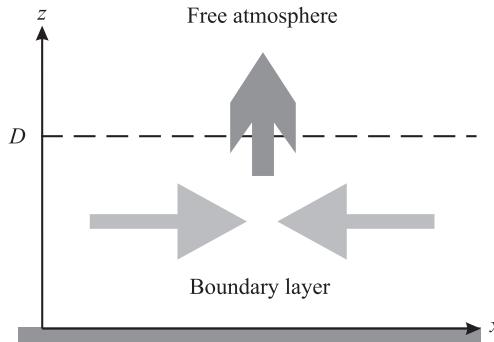


Fig. 5.8

A schematic illustration of the frictional boundary layer, with the free atmosphere above. The dashed line indicates the notional top of the boundary layer. Note that horizontal convergence in the boundary layer (indicated by the lightly shaded arrows) is balanced by flow out of the top of the boundary layer (indicated by the darker arrow). This is the Ekman pumping or suction effect.

the lowest layer of the atmosphere. If the flow is steady, with $\partial/\partial t = 0$, then equations (5.46) give

$$(U_\tau, V_\tau, 0) = \frac{1}{\rho_0 f_0} \left(-\tau_0^{(y)}, \tau_0^{(x)}, 0 \right) = \frac{1}{\rho_0 f_0} \mathbf{k} \times \boldsymbol{\tau}_0, \quad (5.47)$$

where $\mathbf{k} = (0, 0, 1)$ is the unit vertical vector, and hence show that the Ekman volume transport in the boundary layer is *perpendicular* to the surface stress. It is important to note that this result does not depend on the details of the vertical variation of the frictional stresses (other than that they vanish above the boundary layer) or on any ‘eddy viscosity’ assumption.

Turning now to the pressure-driven flow, equations (5.44) show that, when this flow is steady, it must also be geostrophic:

$$u_p = -\frac{1}{\rho_0 f_0} \frac{\partial p}{\partial y}, \quad v_p = \frac{1}{\rho_0 f_0} \frac{\partial p}{\partial x};$$

cf. equations (4.26a) and (4.26b). Then the incompressibility condition (5.15c) implies that the vertical velocity depends only on the stress-driven flow:

$$\frac{\partial w}{\partial z} = - \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = - \left(\frac{\partial u_\tau}{\partial x} + \frac{\partial v_\tau}{\partial y} \right). \quad (5.48)$$

Integration of equation (5.48) through the depth of the boundary layer yields

$$[w]_0^D = - \left(\frac{\partial U_\tau}{\partial x} + \frac{\partial V_\tau}{\partial y} \right); \quad (5.49)$$

the term on the right-hand side of this equation represents the horizontal convergence of the Ekman volume transport. The ground is flat, so $w = 0$ at $z = 0$, and, since the fluid is assumed incompressible, the horizontal convergence in the boundary layer must be balanced by an upward flow w_D out of the top of the boundary layer, at $z = D$; see Figure 5.8. Conversely, horizontal divergence in the boundary layer is balanced by a

downward flow into the top of the boundary layer. Using equations (5.47) and (5.49) this vertical flow can be written in terms of the surface stress component:

$$w_D = -\frac{1}{\rho_0 f_0} \left(\frac{\partial \tau_0^{(x)}}{\partial y} - \frac{\partial \tau_0^{(y)}}{\partial x} \right). \quad (5.50)$$

The velocity w_D is called the **Ekman pumping** or **suction** velocity and equation (5.50) shows that it is proportional to the ‘curl’ of the surface stress.

5.6.2 The laminar Ekman layer

The theory so far has given information only on the vertically integrated properties of the boundary layer. To look at the vertical structure it is necessary to make an assumption about the relationship between the frictional stress τ and the velocity. A simple relationship of this kind is to assume that the stress components are proportional to the vertical gradients of the horizontal velocity:

$$\tau^{(x)} = \rho_0 v \frac{\partial u}{\partial z}, \quad \tau^{(y)} = \rho_0 v \frac{\partial v}{\partial z}. \quad (5.51)$$

This is clearly analogous to the molecular viscous stress law (4.13). The quantity v is the **kinematic eddy viscosity** and corresponds to the viscosity coefficient η , defined in equation (4.13), divided by the density. As noted above, we must use the eddy-viscosity concept with caution.

Using equations (5.51) we now find the vertical structure of the flow within the boundary layer, again assuming that the flow is steady. In principle the viscosity v could vary with z , but we take it to be constant for simplicity. We take $f_0 > 0$, corresponding to the Northern Hemisphere.

Under these conditions the pressure-driven flow (u_p, v_p) is again geostrophic; since the boundary layer is shallow, we take (u_p, v_p) to be independent of z within the boundary layer and equal to the large-scale flow (also assumed geostrophic) just above the boundary layer. For simplicity, we take this large-scale flow to be purely zonal, but varying with the northward distance y ; thus

$$u_p = U(y), \quad v_p = 0, \quad (5.52)$$

say. (The calculation can be re-worked with $v_p \neq 0$; see Problem 5.8.) The boundary conditions on the total flow are

$$(u, v) \rightarrow (0, 0) \quad \text{as } z \rightarrow 0,$$

since, with friction, there can be no flow at the ground, and

$$(u, v) \rightarrow (U, 0) \quad \text{for } z \gg D,$$

assuming that the flow tends to the geostrophic, pressure-driven, value far above the boundary layer. In terms of the stress-driven flow, these become

$$(u_\tau, v_\tau) \rightarrow (-U, 0) \quad \text{as } z \rightarrow 0, \quad (5.53)$$

$$(u_\tau, v_\tau) \rightarrow (0, 0) \quad \text{for } z \gg D, \quad (5.54)$$

using equations (5.43). From equations (5.45), assuming steady flow and using the stress law (5.51), the stress-driven flow satisfies the coupled differential equations

$$-f_0 v_\tau = v \frac{d^2 u_\tau}{dz^2}, \quad f_0 u_\tau = v \frac{d^2 v_\tau}{dz^2}. \quad (5.55)$$

These can be reduced to a single fourth-order differential equation for u_τ or v_τ , but are more easily solved by introducing the complex quantity

$$\lambda_\tau \equiv u_\tau + i v_\tau,$$

which, from equations (5.55), satisfies the second-order equation

$$\frac{d^2 \lambda_\tau}{dz^2} = \frac{i f_0}{v} \lambda_\tau.$$

The two solutions of this are

$$\lambda_\tau \propto \exp \left\{ \pm \left[\frac{i f_0}{v} \right]^{1/2} z \right\} = \exp \left\{ \pm \left[(1+i) \left(\frac{f_0}{2v} \right)^{1/2} \right] z \right\}. \quad (5.56)$$

To satisfy the boundary condition (5.54) we must choose the minus sign in equation (5.56), to ensure exponential decay with z . Then, taking the real and imaginary parts of λ_τ , applying the boundary condition (5.53) and adding the pressure-driven flow (5.52), we get the full solution

$$u = U(y) [1 - e^{-z/h} \cos(z/h)], \quad (5.57a)$$

$$v = U(y) e^{-z/h} \sin(z/h), \quad (5.57b)$$

where

$$h \equiv \left(\frac{2v}{f_0} \right)^{1/2}.$$

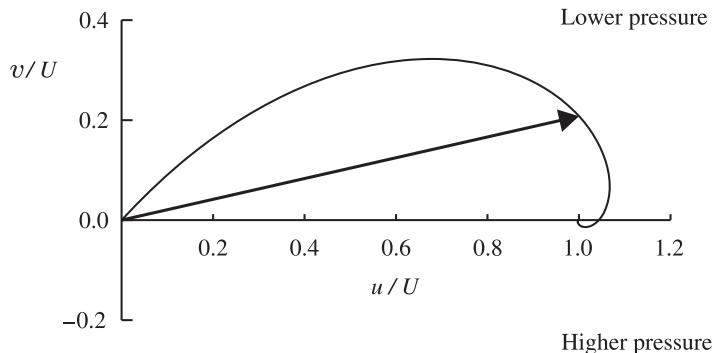
This is **Ekman's solution** and the corresponding boundary layer is called the **Ekman layer**.

We can calculate the corresponding Ekman pumping velocity (5.50) using the stress law (5.51) and equations (5.57):

$$w_D = \left(\frac{v}{2f_0} \right)^{1/2} \left(-\frac{dU}{dy} \right) = \frac{1}{2} h \xi, \quad (5.58)$$

where $\xi = -dU/dy$ is the relative vorticity of the free-atmosphere flow. This relationship between w_D and ξ also holds when the free-atmosphere flow is not purely zonal. Equation (5.58) shows, for example, that the Ekman pumping velocity is upwards under a cyclone in the free atmosphere (for which $\xi > 0$) and downwards under an anticyclone (for which $\xi < 0$).

Figure 5.9 gives a graphical representation (a hodograph) of the horizontal velocity vector (u, v) as a function of z/h . This diagram is called the **Ekman spiral**; note how the velocity vector spirals in towards the free-atmosphere flow $(U, 0)$. It is conventional to define the 'depth' of the Ekman layer to be the height at which the horizontal flow first becomes parallel to the free-atmosphere flow; this corresponds to $z = \pi h = \pi(2v/f_0)^{1/2}$.

**Fig. 5.9**

The Ekman spiral: a hodograph of horizontal wind components (normalised by U) in the Ekman layer. The horizontal axis gives the direction of the geostrophic wind above the boundary layer, so the pressure decreases along the vertical axis. The diagonal arrow indicates the wind vector at a height $z = h\pi/2$.

Putting this equal to 1 km, the approximate depth of the atmospheric boundary layer, allows us to get an order-of-magnitude estimate of the value of the eddy viscosity as $\nu \sim 5 \text{ m}^2 \text{ s}^{-1}$. (This should be contrasted with the kinematic molecular viscosity of air at STP, which is a factor of about 10^{-5} smaller.)

The Ekman spiral also shows that the deflection of the wind in the boundary layer is mostly to the low-pressure side of the geostrophic, large-scale flow. Recall from Section 4.8 that the geostrophic flow blows along the isobars; the Ekman-layer analysis indicates how this is altered in the presence of friction. However, it should be noted that the assumptions that have been made in this calculation are seldom fully satisfied in the atmosphere, owing to the presence of temporal variations and, probably, to the inappropriateness of the eddy-viscosity concept. As a result, a pure Ekman spiral is hardly ever observed in the real atmospheric boundary layer.

5.7 Instability

In Section 2.5 we noted that a compressible atmosphere is statically stable if the squared buoyancy frequency N^2 is positive and statically unstable if it is negative. A similar criterion, but involving N_B^2 rather than N^2 , applies to an incompressible atmosphere under the Boussinesq approximation. Static instability is probably the simplest type of fluid instability to appreciate: at the most basic level it occurs if a layer of dense, incompressible fluid of density ρ_2 is introduced over a layer of lighter incompressible fluid of density $\rho_1 < \rho_2$ in a closed container (Figure 5.10). In this case the dense fluid will fall through the lighter fluid, perhaps in a complicated way, leading eventually to a statically stable state in which the dense fluid is entirely below the light fluid.

A more complex example would be a statically unstable region of finite depth in a compressible atmosphere. Here parcels displaced adiabatically upwards are ‘lighter’

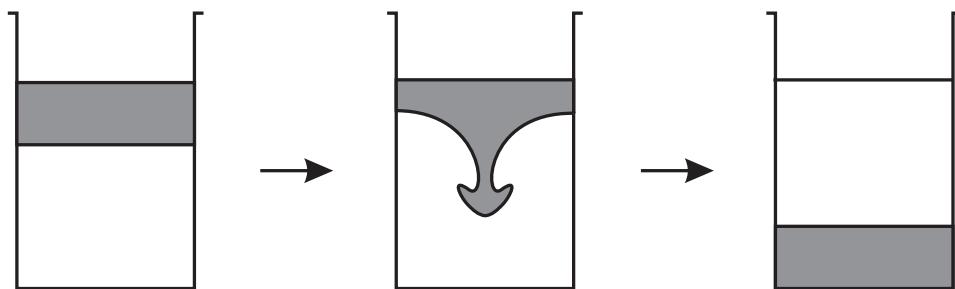


Fig. 5.10

A schematic illustration of static instability in an incompressible fluid. In the left-hand panel, dense fluid (shaded) is introduced over light fluid (unshaded). Static instability leads to a redistribution of the fluids; in the final stable state, shown in the right-hand panel, the dense fluid is entirely below the light fluid.

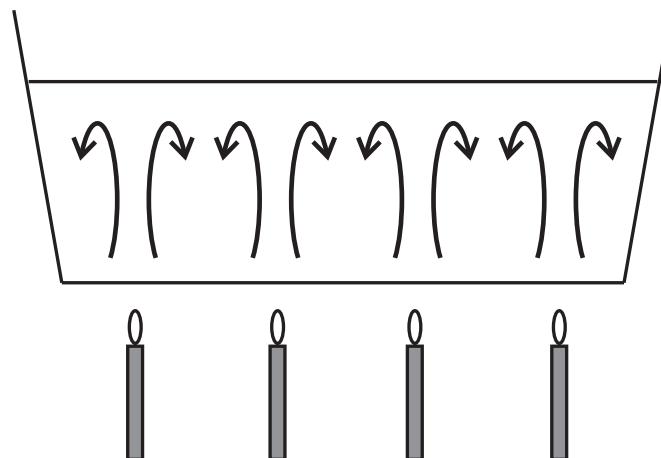


Fig. 5.11

A schematic illustration of convection patterns in a fluid heated from below and cooled from above.

(i.e. have a higher potential temperature θ) than their surroundings and continue to rise, whereas parcels displaced downwards are ‘heavier’ (i.e. have a lower θ) than their surroundings and continue to fall. Eventually a statically stable state is reached, in which low values of θ are at the bottom and higher values are at the top, so that $N^2 = (g/\theta)(d\theta/dz) > 0$.

A third case is one in which local regions of static instability are set up by heating from below, leading to **convective instability**: see Figure 5.11 for a laboratory analogue. In the presence of heating of this kind in the atmosphere we may expect much of the forced region to approach neutral stability, with N^2 approaching zero and the lapse rate approaching the dry adiabatic value Γ_a (see Section 2.5) or the saturated adiabatic value Γ_s (see Section 2.8) according to whether the air is unsaturated or saturated.

On large scales, the rotation of the Earth allows more subtle types of instability, which can lead to important atmospheric disturbances such as cyclones and anticyclones and other weather phenomena. We briefly discuss the two most important of these: **baroclinic instability** and **barotropic instability**.

5.7.1 Baroclinic instability

Consider a steady zonal flow $(U(z), 0, 0)$ in the Northern Hemisphere that increases with height and work with the Boussinesq equations on an f -plane. From geostrophic balance (5.18) and hydrostatic balance (5.20) we can derive a thermal windshear equation (cf. equation (4.28b))

$$f_0 \frac{dU}{dz} = \frac{g}{\rho_0} \frac{\partial \rho}{\partial y},$$

where $\rho = \bar{\rho}(z) + \rho'$. Since $dU/dz > 0$ and $f_0 > 0$ this implies that $\partial \rho / \partial y > 0$, i.e. the density increases with northward distance. Now the slope α of the density surfaces in the y, z plane satisfies

$$\tan \alpha \equiv \left(\frac{\partial z}{\partial y} \right)_\rho = -\frac{\partial \rho / \partial y}{\partial \rho / \partial z}$$

by the reciprocity theorem of partial differentiation, equation (4.30). Assuming that the flow is statically stable, so that $\partial \rho / \partial z < 0$, it follows that $\tan \alpha > 0$. Thus a zonal wind that increases with height is associated, through thermal windshear balance, with density surfaces that slope polewards and upwards.

We now show that this configuration of density surfaces may be able to release potential energy. Suppose that air parcels, each of volume V , are somehow interchanged between two density surfaces $\rho = \rho_1$ and $\rho = \rho_2$, where $\rho_2 > \rho_1$, as shown in Figure 5.12. That is, parcel A, of mass $m_1 = \rho_1 V$ and at initial height z_1 , is swapped with parcel B, of mass $m_2 = \rho_2 V$ and initial height z_2 , without any other part of the fluid being disturbed. The initial potential energy of the two parcels alone is $g(m_1 z_1 + m_2 z_2)$ while their final potential energy is $g(m_2 z_1 + m_1 z_2)$. There is therefore an increase of potential energy of the two parcels (and hence of the whole fluid, since no other changes take place) given by

$$\begin{aligned} \Delta E_P &= g(m_2 z_1 + m_1 z_2) - g(m_1 z_1 + m_2 z_2) \\ &= -g(m_2 - m_1)(z_2 - z_1) = -gV(\rho_2 - \rho_1)(z_2 - z_1). \end{aligned}$$

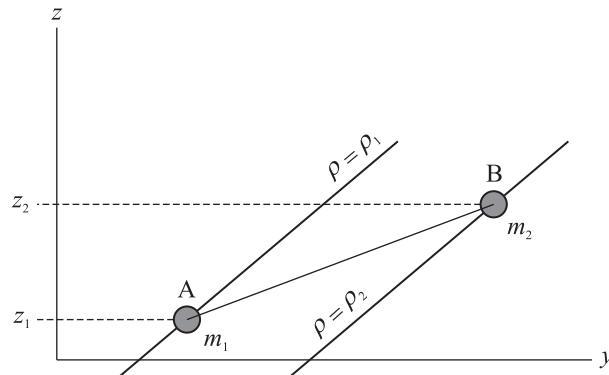


Fig. 5.12

Hypothetical interchange of two parcels from surfaces of different densities.

However, $\rho_2 > \rho_1$, so, if $z_2 > z_1$, as in Figure 5.12 (that is, the heavier parcel is initially higher than the lighter parcel), then $\Delta E_p < 0$; there is thus a decrease of potential energy as a result of the interchange. This can occur, despite the stable density stratification, because the density surfaces are sloping more steeply than the line joining the two parcels. There is then a possibility that the potential energy released by the interchange may be converted to kinetic energy, leading to unstable motions. These motions are an example of **baroclinic instability**, also known as **sloping convection**.⁶ Of course this kind of interchange of parcels is highly idealised, so further analysis is needed to determine whether it is ever dynamically possible.

A simple model, which is more realistic and also dynamically consistent, is **Eady's model** of baroclinic instability, in which the background flow is taken to be a linear function of height, $U = \Lambda z$, between rigid horizontal boundaries at $z = D$ and $z = -D$ on an f -plane, with N_B constant. The corresponding background streamfunction is $\Psi = -\Lambda yz$ and the background density is a linear function of y .

Linearised quasi-geostrophic disturbances to this flow are considered, and for simplicity we shall assume that these disturbances are independent of y : they therefore satisfy a linearised QGPV equation of the form

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \left(\frac{\partial^2 \psi'}{\partial x^2} + \frac{f_0^2}{N_B^2} \frac{\partial^2 \psi'}{\partial z^2} \right) = 0, \quad (5.59)$$

(cf. equations (5.25) and (5.35)). The boundary conditions are that the vertical velocity $w_a = 0$ at $z = -D$ and D ; from equation (5.24) this implies that $D_g(\partial \psi / \partial z) = 0$ at the boundaries. On linearisation about the background flow this gives

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \frac{\partial \psi'}{\partial z} - \Lambda \frac{\partial \psi'}{\partial x} = 0 \quad \text{at } z = \pm D. \quad (5.60)$$

Normal mode solutions of equations (5.59) and (5.60) of the form

$$\psi' = \text{Re } \hat{\psi}(z) \exp[ik(x - ct)] \quad (5.61)$$

are now sought, where $k > 0$ and the phase speed c may be complex: $c = c_r + ic_i$. It is found that $c_i > 0$ for a certain range of wavenumbers, corresponding to a normal mode that grows in time like $\exp(kc_i t)$: this exponential growth is a signal of instability. The quantity kc_i is called the **growth rate**; it is the inverse of the e -folding time for the instability. A brief sketch of the details is as follows.

Substitution of (5.61) into (5.59) gives

$$ik(U - c) \left(-k^2 \hat{\psi} + \frac{f_0^2}{N_B^2} \hat{\psi}'' \right) = 0$$

⁶ The word **baroclinic** in this context refers to the height variation of the zonal wind or the corresponding latitudinal variation of the density or temperature.

and hence, if $U(z) \neq c$,

$$\hat{\psi}'' - K^2 \hat{\psi} = 0, \quad \text{where } K = \frac{N_B k}{f_0};$$

note that this corresponds to a perturbation QGPV that is zero everywhere. The solutions can be written

$$\hat{\psi} = A \cosh(Kz) + B \sinh(Kz), \quad (5.62)$$

where A and B are complex constants. Substitution of (5.61) into the boundary conditions (5.60) gives

$$\left(z - \frac{c}{\Lambda}\right) \hat{\psi}' - \hat{\psi} = 0 \quad \text{at } z = \pm D.$$

When expression (5.62) is substituted into each of these boundary conditions, the resulting equations added and subtracted and then A and B eliminated, we obtain the dispersion relation

$$c^2 = -\frac{\Lambda^2}{K^2} [\coth(KD) - KD][KD - \tanh(KD)]. \quad (5.63)$$

It can be shown that the right-hand side of equation (5.63) is negative for $0 < KD \lesssim 1.2$ and positive otherwise. Therefore we get growing modes ($c_i > 0$) for non-zero wavenumbers k less than about $1.2f_0/(N_B D)$, i.e. horizontal wavelengths greater than about $5.2N_B D/f_0$; these unstable modes have $c_r = 0$ and do not propagate. The maximum growth rate kc_i is about $0.31\Lambda f_0/N_B$ and occurs where $k \approx 0.8f_0/(N_B D)$: see Figure 5.13. For $k \gtrsim 1.2f_0/(N_B D)$, c is real and no growth occurs; the resulting stable modes propagate with equal and opposite real phase speeds c_r .

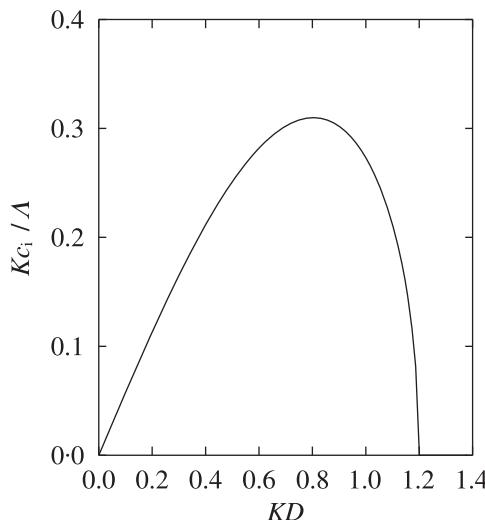


Fig. 5.13

Normalised growth rate Kc_i/Λ as a function of KD for unstable normal modes in Eady's model of baroclinic instability. Note that the ordinate is proportional to the growth rate kc_i and the abscissa is proportional to the horizontal wavenumber k .

Although the Eady model is highly simplified, it can be generalised in various ways. The unstable modes that result bear some resemblance to developing cyclones and anticyclones in the midlatitude troposphere and also to certain unstable waves observed in laboratory analogues of the atmosphere (see Section 9.4).

5.7.2 Barotropic instability

A different type of instability can occur for basic winds $U(y)$ that vary with the northward distance y but not with height z ; this kind of flow structure is called **barotropic**. Taking the disturbance streamfunction ψ' to be independent of z and retaining the β -effect, we get

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \left(\frac{\partial^2 \psi'}{\partial x^2} + \frac{\partial^2 \psi'}{\partial y^2} \right) + (\beta - U'') \frac{\partial \psi'}{\partial x} = 0,$$

from linearising the QGPV equation (5.25). We seek normal mode solutions of the form

$$\psi = \text{Re } \hat{\phi}(y) \exp[ik(x - ct)], \quad (5.64)$$

where again $k > 0$ and c may be complex, and obtain the second-order ordinary differential equation

$$(U - c)(\hat{\phi}'' - k^2 \hat{\phi}) + (\beta - U'')\hat{\phi} = 0. \quad (5.65)$$

This must be solved subject to suitable boundary conditions; convenient (though not physically very realistic) ones are that the northward velocity should vanish on east–west oriented boundaries at $y = 0, L$, say. Hence

$$\hat{\phi} = 0 \quad \text{at } y = 0, L. \quad (5.66)$$

Equations (5.65) and (5.66) form a system with complex eigenvalues $c^{(n)} = c_r^{(n)} + i c_i^{(n)}$, $n = 1, 2, 3, \dots$ and corresponding complex eigenfunctions $\hat{\phi}^{(n)}(y)$. Unlike the Eady-problem case, these equations cannot generally be solved analytically; however, they are straightforward to solve by numerical methods.

The physical interpretation of these solutions is as follows. Suppose that the zonal flow $U(y)$ is initially undisturbed, for $t < 0$. At time $t = 0$ it is given a small perturbation ψ_0 ; this perturbation can be Fourier-analysed into components of differing wavenumber k ; moreover each of these components can in turn be resolved into a linear combination of the eigenfunctions for that value of k . Each term (or mode) in this sum then evolves in time according to equation (5.64), with its own complex phase speed. Modes for which $c_i \leq 0$ are stable: their amplitudes do not grow with time. However, any modes for which $c_i > 0$ will grow exponentially in time like $\exp(kc_i t)$; these are called unstable modes and the mode with the largest c_i will eventually outgrow the others and dominate the disturbance. Thus if unstable modes exist, most initial disturbances will generally excite them and the flow will rapidly be distorted. (Of course, the exponential growth in this case means that the assumption of a small disturbance amplitude will eventually break down for sufficiently large t , whereupon the model ceases to be valid.)

This **barotropic instability** mechanism is believed to be responsible for a number of types of large-scale weather disturbance observed in the tropical troposphere, as well as eddies in the stratosphere. On the other hand, if there are no unstable modes, initial disturbances will not grow and the flow is not affected very much by the perturbation imposed at $t = 0$.

A useful general result is the **Rayleigh–Kuo criterion**, which states that if $\beta - U''$ does not change sign in the region $0 \leq y \leq L$ then all modes are stable. It does not guarantee that unstable modes will always occur if $\beta - U''$ vanishes somewhere; however, the latter condition must hold if there is to be any chance of finding unstable modes.

Further reading

The physical interpretation of vorticity given here follows the approach of Acheson (1990). Vector-calculus identities like equation (5.3) are given in many mathematics texts, such as those by Riley *et al.* (2006) and Boas (1983). The advanced texts on geophysical fluid dynamics by Vallis (2006), Pedlosky (1987) and Gill (1982) cover most of the topics in this chapter in much more detail than attempted here, and should be consulted by readers who wish to pursue the subject further. The original derivation of equation (5.40) was given in the pioneering paper by Charney and Drazin (1961).

Problems

Problem 5.1 Show that the two-dimensional circular flow $\mathbf{u} = V(r)\mathbf{i}_\phi$ introduced in Section 5.1 can be written in plane Cartesian coordinates (x, y, z) as

$$\mathbf{u} = \frac{V}{(x^2 + y^2)^{1/2}}(-y, x, 0);$$

hence or otherwise verify that the z component of the vorticity is

$$\xi = \frac{dV}{dr} + \frac{V}{r},$$

and that the other components are zero.

Problem 5.2 Pure inertial oscillations (see also Problem 4.3) can be modelled by neglecting horizontal pressure gradients in the linearised Boussinesq equations (5.16a) and (5.16b) on an f -plane:

$$u_t - f_0 v = 0, \quad v_t + f_0 u = 0.$$

Define $\tilde{u} = u + iv$ and solve for (u, v) , given that $(u, v) = (u_0, 0)$ at $t = 0$. Given that **particle displacements** (X, Y) satisfy

$$\frac{\partial X}{\partial t} = u, \quad \frac{\partial Y}{\partial t} = v,$$

show that the particle trajectories for these oscillations are circular. Calculate the radius of the circle if the particle speed is (a) 1 m s^{-1} and (b) 10 m s^{-1} .

Problem 5.3 Calculate the horizontal and vertical air-parcel displacements X and Z associated with an internal gravity wave, defined by $X_t = u$ and $Z_t = W$ (cf. Problem 5.2), given that these displacements vanish when $x = z = t = 0$. Show that the air parcels oscillate in straight lines perpendicular to the vector $(k, 0, m)$.

Find the period (in minutes) of an internal gravity wave of horizontal wavelength 100 km and vertical wavelength 5 km in the Earth's mesosphere, where $N_B^2 = 3 \times 10^{-4} \text{ s}^{-2}$. How long (in minutes) does the information associated with this wave take to propagate through a vertical distance of 20 km? If the maximum horizontal wind fluctuation (peak-to-peak) due to the wave is 2 m s^{-1} , find the maximum horizontal and vertical distances traversed by an air parcel.

Problem 5.4 **Inertia–gravity waves** are the generalisation of the internal gravity waves of Section 5.4 to the case when $f_0 \neq 0$. Look for linear plane-wave solutions of the form (5.27) and show in particular that

$$\rho_0 \hat{u} = \frac{\omega k \hat{p}}{\omega^2 - f_0^2}, \quad \rho_0 \hat{v} = \frac{-ikf_0 \hat{p}}{\omega^2 - f_0^2} \quad (5.67)$$

and that the dispersion relation is

$$\omega^2 = f_0^2 + \frac{N_B^2 k^2}{m^2}.$$

Note that the northward velocity v must be non-zero.

What is the minimum angular frequency of these waves? Show that, for a given frequency and vertical wavelength, these waves have a larger horizontal wavelength than do the corresponding internal gravity waves.

Problem 5.5 Using equations (5.67) and the hydrostatic equation, calculate the mean kinetic energy per unit volume \bar{K} and the mean available potential energy per unit volume \bar{P} , averaged over one wave period. How does the ratio \bar{P}/\bar{K} behave when $\omega \rightarrow f_0$ and when $\omega \gg f_0$? To what kind of wave does the latter limit refer? Show that equipartition between mean kinetic energy and mean available potential energy does not occur for long waves.

Problem 5.6 Starting with the Rossby-wave dispersion relation (5.36), consider waves in the absence of a background flow, with $l = 0$. For fixed m put $b = f_0 m / N_B$ and sketch ω as a function of k , for $k > 0$, with particular attention to the limits $k \rightarrow 0$ and $k \gg b$. Show that Rossby waves cannot exist if the wave period is shorter than a critical value. Estimate this critical period at 45° N for waves of vertical wavelength 10 km, given $N_B^2 = 5 \times 10^{-4} \text{ s}^{-2}$, a value representative of the stratosphere.

Problem 5.7 Consider Rossby waves that vary with x and y but are independent of depth z , in a uniform zonal flow U . Show that stationary waves can exist only if the zonal (x)

wavelength is greater than a certain value, L say. Calculate L at 60° N for (a) $U = 50 \text{ m s}^{-1}$ and (b) $U = 150 \text{ m s}^{-1}$. Compare this with the length of a latitude circle and comment.

Problem 5.8 Take $(u, v) = (u_p + u_\tau, v_p + v_\tau)$, where $u_p(x, y)$, $v_p(x, y)$ is a geostrophic flow and $u_\tau(z)$ and $v_\tau(z)$ satisfy the steady laminar Ekman-layer equations

$$-fv_\tau = \nu \frac{\partial^2 u_\tau}{\partial z^2}, \quad fu_\tau = \nu \frac{\partial^2 v_\tau}{\partial z^2},$$

(f is a constant, taken to be positive, for the Northern Hemisphere). The boundary conditions are $(u, v) = 0$ on $z = 0$ and $(u, v) \rightarrow (u_p, v_p)$ as $\zeta \rightarrow \infty$, where $\zeta = z/h$ and $h = (2\nu/f)^{1/2}$.

Put $\lambda_\tau = u_\tau + iv_\tau$ and show that

$$\lambda_\tau = -(u_p + iv_p)e^{-(1+i)\zeta}. \quad (5.68)$$

Hence verify that

$$\begin{aligned} u &= u_p(1 - e^{-\zeta} \cos \zeta) - v_p e^{-\zeta} \sin \zeta, \\ v &= v_p(1 - e^{-\zeta} \cos \zeta) + u_p e^{-\zeta} \sin \zeta. \end{aligned}$$

Choose axes such that $v_p = 0$ and draw the Ekman spiral, showing how the vector (u, v) varies as ζ increases from 0 to ∞ . Note that friction causes a significant cross-isobar flow in the Ekman layer: in which direction is this flow, in general? Sketch the relative directions of the pressure-gradient force, the Coriolis force and the viscous force at some point above the ground but well within the boundary layer, assuming that $u_p > 0$.

Using equation (5.68), calculate $\int_0^\infty u_\tau dz$ and $\int_0^\infty v_\tau dz$. What do these quantities represent? Show that the vertical velocity at the top of the Ekman layer is given by

$$w_\tau = \left(\frac{\nu}{2f}\right)^{1/2} \left(\frac{\partial v_p}{\partial x} - \frac{\partial u_p}{\partial y}\right).$$

If the boundary layer depth is of order 1 km and $f = 10^{-4} \text{ s}^{-1}$, estimate the order of magnitude of ν . If the geostrophic wind changes by 20 m s^{-1} over a horizontal distance of 1000 km, estimate the order of magnitude of w_τ . Give the direction and order of magnitude of the horizontal stress at the ground at a point where the geostrophic wind is 10 m s^{-1} . (Assume that $\rho = 1 \text{ kg m}^{-3}$.)

Problem 5.9 Suppose that friction is represented by linear **Rayleigh friction**

$$F^{(x)} = -ru, \quad F^{(y)} = -rv,$$

rather than by an eddy-viscosity assumption. Use this representation in the linear Boussinesq momentum equations on an f plane to model the following situation.

During the daytime, convection is active, leading to a strong frictional coupling with the ground (modelled by taking r to be large, $r = 10^{-5} \text{ s}^{-1}$) and the atmosphere can be assumed to be in a steady state at this time. At sunset, convection is quenched and r becomes zero. Assuming that the pressure gradient is purely in the y direction, is independent of height and does not change with time, show that the subsequent motion consists of geostrophic

motion, superimposed on which is an inertial oscillation of amplitude $r u_g / (r^2 + f^2)^{1/2}$, where u_g is the geostrophic velocity.

Calculate the amplitude of the oscillation at 30°N when the flow is 20 m s^{-1} from the west. Show that after 12 h the meridional flow will be the reverse of its daytime value. (Assume that night-time conditions last this long.)

Problem 5.10 Consider a region of the atmosphere, in the Northern Hemisphere, that is in geostrophic and hydrostatic balance. Make the Boussinesq approximation and assume that the wind is in the zonal direction and increases linearly with height, while the density ρ decreases linearly with height.

Show that the surfaces of constant density slope upwards towards the pole at an angle α to the horizontal given by

$$\tan \alpha = f \Lambda / N_B^2,$$

where Λ is the vertical wind shear and N_B is the buoyancy (Brunt–Väisälä) frequency.

Now suppose that two air parcels of equal volume V , densities ρ_1 and ρ_2 (where $\rho_2 > \rho_1$) and heights z_1 and z_2 are interchanged, while the rest of the atmosphere remains undisturbed. Under what condition is potential energy released by this process?

Suppose that $\delta\rho = \rho_2 - \rho_1$ and the distance δs between the parcels are small. Show that the amount of potential energy released is given by

$$\Delta P = \rho_0 V (N_B \delta s)^2 (\sin \phi \cos \phi \tan \alpha - \sin^2 \phi),$$

where ϕ is the angle between the line joining the parcels and the horizontal. [Hint: note that $\delta\rho = (\partial\rho/\partial y) \delta y + (\partial\rho/\partial z) \delta z$, where $\delta y = \delta s \cos \phi$, $\delta z = \delta s \sin \phi$.]

Show that ΔP reaches a maximum ΔP_{\max} when $\phi = \alpha/2$. Given that α is small, so that $\alpha \approx \tan \alpha$, show that

$$\Delta P_{\max} \approx \rho_0 V \left(\frac{f \Lambda \delta s}{2 N_B} \right)^2.$$

Estimate the angle to the horizontal at which air parcels would move to maximise the release of potential energy near 60°N , given a vertical windshear of $2 \text{ m s}^{-1} \text{ km}^{-1}$ and a buoyancy period of 8 min.

Problem 5.11 Apply the Eady model of baroclinic instability to a region of the atmosphere near 50°N in which the mean zonal wind varies by 30 m s^{-1} over a depth of 10 km and $N_B \simeq 10^{-2} \text{ s}^{-1}$. Estimate the time (in days) taken by the fastest-growing Eady mode to grow in amplitude by a factor of e and estimate the zonal wavelength of this mode. What is the shortest zonal wavelength for which instability can occur under the given conditions?

In keeping with the emphasis on atmospheric physics in this book, the purpose of the present chapter is to illustrate the use of basic physical principles in the study of some aspects of atmospheric chemistry, rather than to provide a comprehensive treatment of atmospheric chemistry as a whole. We therefore focus on stratospheric chemistry, which provides some simple yet important applications of the basic principles and also some examples of interactions between chemistry and dynamics.

In [Section 6.1](#) we outline some of the basic thermodynamics of chemical reactions, while in [Section 6.2](#) we introduce some elementary aspects of chemical kinetics, including the concepts of reaction rates and chemical lifetimes. In [Section 6.3](#) we focus on bimolecular reactions and show how physical reasoning can give an expression for the reaction rate. The process of photo-dissociation is introduced in [Section 6.4](#). Once these basic ideas have been established, we apply them to stratospheric ozone in [Section 6.5](#), first describing the Chapman theory (which involves oxygen compounds only) and then introducing the effects of catalytic cycles. The principles of chemical transport by atmospheric flows are discussed in [Section 6.6](#), with a qualitative description of the main global-scale meridional transport structures in the middle atmosphere. Finally, in [Section 6.7](#), we bring several of these ideas together in a general description of the processes implicated in the formation of the Antarctic ozone hole.

6.1 Thermodynamics of chemical reactions

Consider a chemical reaction in which reactants A and B lead to products C and D:



In the laboratory this reaction takes place within a fixed reaction vessel; in the atmosphere we have to imagine it taking place within a given ‘parcel’ or ‘blob’ of air, with no transfer of mass into or out of the parcel. If the reaction takes place at constant pressure p , which is usually true for reactions in the atmosphere, then it follows from the First Law of Thermodynamics in the form [\(2.18\)](#) and from [equation \(2.19\)](#) that the heat ΔQ supplied to the parcel is given, assuming reversibility, by

$$\Delta Q = \Delta H.$$

Here ΔH is the **enthalpy of reaction**, that is, the sum of the enthalpies of the products minus the sum of the enthalpies of the reactants. If $\Delta H > 0$, so that heat must be supplied for the reaction to proceed, the reaction is called **endothermic**, whereas if $\Delta H < 0$, so that heat is liberated, the reaction is called **exothermic**.

The **standard molar enthalpy of formation** of a compound is the enthalpy of reaction associated with the formation of one mole of the compound from its constituent elements at a standard temperature (e.g. $25^\circ\text{C} = 298.15\text{ K}$) and pressure (1 atm). This is often denoted by ΔH_f^\ominus , the superscript \ominus referring to the standard temperature and pressure. Tables of ΔH_f^\ominus for a variety of atmospheric gases are available in atmospheric chemistry texts. (The corrections required to convert the enthalpy of reaction to other atmospheric temperatures and pressures are fairly small.)

Applying these ideas to reaction (6.1) we obtain, at standard temperature and pressure (STP), the molar enthalpy of reaction

$$\Delta H^\ominus = \Delta H_f^\ominus(\text{C}) + \Delta H_f^\ominus(\text{D}) - \Delta H_f^\ominus(\text{A}) - \Delta H_f^\ominus(\text{B}). \quad (6.2)$$

Given the enthalpies of formation on the right-hand side of this equation, we can determine whether the reaction is endothermic or exothermic. As an example, consider the recombination of oxygen (O) atoms to molecular oxygen (O_2) near the mesopause,



Here M is an arbitrary air molecule, which is required to satisfy conservation of energy and momentum in the reaction. Typically M is either of the most abundant atmospheric molecules, N_2 or O_2 . Reaction (6.3) is found to have a negative enthalpy of reaction $\Delta H \approx -500\text{ kJ mol}^{-1}$, and is thus exothermic; it therefore contributes to the diabatic heating of the atmosphere; see Section 4.10.

Conversely, the photolysis (or photo-dissociation) of O_2 to form two O atoms has $\Delta H \approx +500\text{ kJ mol}^{-1}$ (the same enthalpies of formation apply as those in reaction (6.3)), so this reaction is endothermic: some source of energy is required to drive it. One possibility, which actually occurs in the atmosphere, is that this energy is supplied by solar photons. A simple calculation (see Problem 6.1) shows that the relevant photons must have wavelengths of less than about 240 nm. The standard notation for such a reaction is



$h\nu$ here referring to the photon energy required for photolysis.

Thermodynamics also allows us to predict the direction in which a reaction will proceed; for example, whether $\text{A} + \text{B} \rightarrow \text{C} + \text{D}$ or $\text{C} + \text{D} \rightarrow \text{A} + \text{B}$. Consider a natural (or spontaneous) change for a ‘system’ (here taken to be our parcel of reacting gases) immersed in a ‘heat bath’ (the surrounding atmosphere) at fixed temperature and pressure. The Second Law of Thermodynamics implies that the Gibbs free energy $G = U + pV - TS$ (see Section 2.10)

of the system (excluding the surroundings) must decrease during this change;¹ i.e.,

$$\Delta G < 0. \quad (6.5)$$

An equation analogous to [equation \(6.2\)](#) relates the overall **molar Gibbs free energy of reaction** at standard temperature and pressure, ΔG^\ominus , to the molar Gibbs free energies of formation of the gases taking part in a reaction. Tabulations of the latter then allow us to determine whether ΔG^\ominus is negative (so that the reaction can occur spontaneously at STP), positive (so that it cannot) or zero (so that there is equilibrium between the gases involved).

Since the enthalpy $H = U + pV$ and $T = T_0$ is constant, we have the following relation between the molar Gibbs free energy of reaction ΔG , the molar enthalpy of reaction ΔH and the change in molar entropy of the system ΔS :

$$\Delta G = \Delta H - T_0 \Delta S.$$

6.2 Chemical kinetics

We now consider how rapidly chemical reactions occur, starting with the simplest **first-order** or **unimolecular** reaction, in which one molecule of a single reactant A breaks down to one molecule of product B and one molecule of product C:



Representing the number density of A as $[A]$ molecules per unit volume, we define the **reaction rate** R_A as the rate of decrease of $[A]$ or equivalently the rate of increase of $[B]$ or $[C]$:

$$R_A = -\frac{\partial [A]}{\partial t} = \frac{\partial [B]}{\partial t} = \frac{\partial [C]}{\partial t} \equiv k_A[A]. \quad (6.6)$$

The term on the far right-hand side of this equation defines the **rate coefficient** k_A , which has units of s^{-1} ; its inverse, the **chemical lifetime** τ_A of A, satisfies

$$\tau_A = \frac{1}{k_A} = \frac{[A]}{\left| \frac{\partial [A]}{\partial t} \right|}.$$

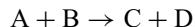
In practice the rate coefficient is independent of $[A]$ and t , so we can integrate [equation \(6.6\)](#) to get

$$[A] = [A]_0 e^{-k_A t} = [A]_0 e^{-t/\tau_A},$$

where the subscript 0 indicates the value at time $t = 0$. This shows that the number density of A decays exponentially in time, decreasing by a factor e in one chemical lifetime.

¹ The proof starts with the statement that the entropy of the system plus the surroundings cannot decrease during the change. Consideration of the heat and work interactions between the system and its surroundings and use of the First Law and the facts that $T = T_0$ and $p = p_0$ are constant during the change then lead to [\(6.5\)](#). Details are given in standard books on thermodynamics, e.g. Blundell and Blundell (2009).

Next we consider **second-order** or **bimolecular** reactions, of the type considered in equation (6.1):



In this case the reaction rate R_A and the rate coefficient k_{AB} satisfy

$$R_A = -\frac{\partial[A]}{\partial t} = -\frac{\partial[B]}{\partial t} = \frac{\partial[C]}{\partial t} = \frac{\partial[D]}{\partial t} \equiv k_{AB}[A][B]. \quad (6.7)$$

We can define an instantaneous lifetime for A for this reaction by

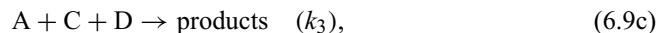
$$\tau_A = \frac{[A]}{\left| \frac{\partial[A]}{\partial t} \right|} = \frac{1}{k_{AB}[B]}.$$

In this case we only get a solution of the form

$$[A] = [A]_0 e^{-t/\tau_A} \quad (6.8)$$

if $[B] \gg [A]$ so that $[B]$ can be considered to be approximately constant in time; see Problem 6.3. Third-order (termolecular) and higher-order reactions can also be considered; some examples are discussed later.

We now consider several simultaneous reactions. First suppose that A has three loss processes, but no production, described by



where the k_i are the rate coefficients. The reaction rate R_A includes contributions from each of these processes. It is given by

$$R_A = -\frac{\partial[A]}{\partial t} = k_1[A] + k_2[A][B] + k_3[A][C][D].$$

The corresponding lifetime for A is

$$\tau_A = \frac{[A]}{\left| \frac{\partial[A]}{\partial t} \right|} = \frac{1}{k_1 + k_2[B] + k_3[C][D]}$$

and $[A]$ decays exponentially with time if $[B]$, $[C]$ and $[D]$ are approximately constant. (The lifetime may still give an instantaneous measure of the fractional rate of change of $[A]$ even if $[B]$, $[C]$ and $[D]$ are varying.)

Now suppose that, in addition to the loss processes described by reactions (6.9), there is also production of A, given by



in this case the reaction rate is

$$R_A = -\frac{\partial[A]}{\partial t} = k_1[A] + k_2[A][B] + k_3[A][C][D] - k_4[E][F].$$

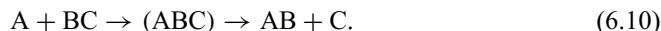
This shows that a steady state is possible, in which $\partial[A]/\partial t = 0$, with a balance between production and loss of A and an equilibrium number density of A given by

$$[A] = \frac{k_4[E][F]}{k_1 + k_2[B] + k_3[C][D]}.$$

It should be emphasised that the equations considered in this section apply to a fixed volume of reacting chemicals. They need modification when applied to a moving parcel of air, whose volume may be changing with time; see [Section 6.6](#).

6.3 Bimolecular reactions

Some physical insight into the nature of chemical reactions can be gained by considering a special case, that of bimolecular reactions, which we now represent by



This starts with initial molecules A and BC; BC breaks into fragments B and C and then A, B and C pass through an intermediate ‘transient reaction complex’ ABC, before forming the final molecules AB and C. A schematic diagram of the intermediate complex is given in [Figure 6.1](#): initially the distance r_{BC} between B and C is small and the distance r_{AB} between A and B is large. During the reaction, the situation is reversed, r_{AB} becoming small and r_{BC} becoming large. In the special case in which the angle $\angle ABC$ is fixed, the process can be represented in a two-dimensional contour plot of the potential energy as a function of r_{AB} and r_{BC} ; see [Figure 6.2](#). This shows that, if sufficient energy is available, ABC can proceed up the ‘potential energy valley’ X, associated with $A + BC$, over the ‘saddle point’ or ‘col’ at P and down into the ‘valley’ Y associated with $AB + C$, following the dashed line in [Figure 6.2](#).

This idea is further illustrated in [Figure 6.3](#), a plot of the potential energy against a ‘reaction coordinate’ r , which indicates how far the reaction has proceeded. This diagram is a vertical slice, along the dashed line, through the potential energy surface whose contours are given in [Figure 6.2](#). Reaction (6.10) proceeds from left to right in [Figure 6.3](#): an **activation energy** $E_1 = H_{ABC} - H_{A+BC}$ is required to get from the valley X over the saddle point at P. In the case shown, a larger amount of energy $E_2 = H_{ABC} - H_{AB+C}$ is regained on the further side, in valley Y, so the reaction is exothermic, with an enthalpy of

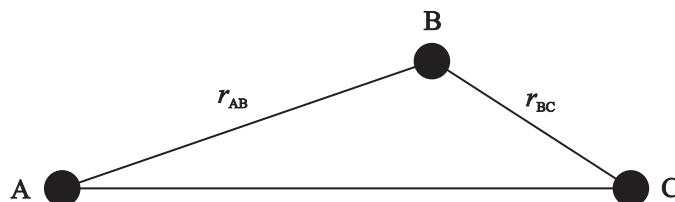


Fig. 6.1

A schematic sketch of the transient reaction complex (ABC) involved in reaction (6.10).

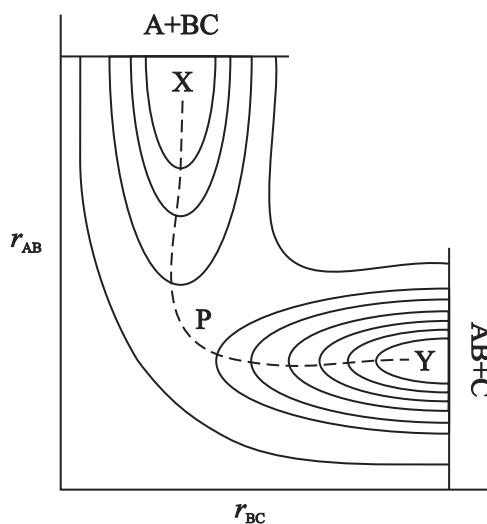


Fig. 6.2 A contour plot of potential energy for reaction (6.10), showing the path of the reaction as a dashed line that proceeds from X, over the saddle point P, to Y.

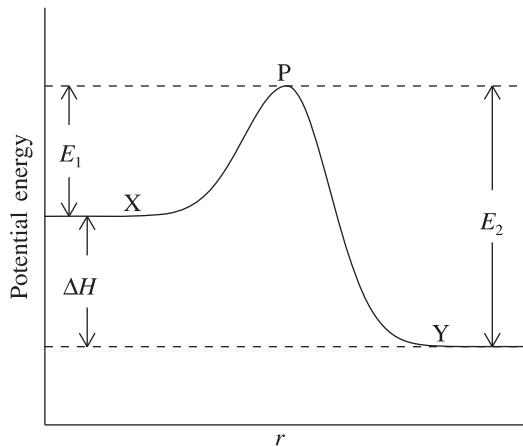


Fig. 6.3 A plot of the potential energy for reaction (6.10), as a function of the reaction coordinate r .

reaction $\Delta H = E_1 - E_2 < 0$. On the other hand, the reverse reaction $AB + C \rightarrow A + BC$, proceeding from right to left in Figure 6.3, is endothermic, with an enthalpy of reaction $\Delta H = E_2 - E_1 > 0$.

A crude **collision theory** provides a simple model for the temperature dependence of the bimolecular rate coefficient. The kinetic theory of gases shows that the number of collisions, per unit volume per unit time, between molecules A and BC is $[A][BC]\sigma\bar{c}$, where σ is the collision cross-section

$$\sigma = \pi(r_A + r_{BC})^2,$$

r_A and r_{BC} being the notional radii of the molecules (taken to be hard spheres in this model). The molecules are taken to have masses m_A and m_{BC} , respectively, and \bar{c} is the mean relative speed of the molecules at temperature T :

$$\bar{c} = \left(\frac{8k_B T}{\pi m_r} \right)^{1/2},$$

where k_B is Boltzmann's constant and m_r is the reduced mass $m_A m_{BC} / (m_A + m_{BC})$ of the molecules. However, not all collisions lead to reaction: the colliding molecules must possess sufficient kinetic energy to overcome the energy barrier E_1 ; the probability of this occurring is given by the Boltzmann factor $\exp[-E_1/(k_B T)]$. The mean rate of loss of molecules A per unit volume is therefore

$$-\frac{\partial [A]}{\partial t} = [A][BC]\pi(r_A + r_{BC})^2 \left(\frac{8k_B T}{\pi m_r} \right)^{1/2} e^{-E_1/(k_B T)} \equiv k_2[A][BC],$$

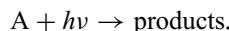
say. Apart from the weak $T^{1/2}$ dependence, the expression for the rate coefficient k_2 is similar to the empirical **Arrhenius expression**

$$k_2 = \alpha e^{-E_a/(k_B T)},$$

where α is a constant and E_a is an activation energy, provided that E_a is associated with the barrier energy E_1 .

6.4 Photo-dissociation

In equation (6.4) we gave an example of photo-dissociation, in which the energy for a chemical reaction is provided by an incident solar photon that is absorbed by the reactant molecule (see Section 3.1). In general we can write such a reaction as



where the number density $[A]$ satisfies

$$\frac{\partial [A]}{\partial t} = -j_A[A],$$

j_A being the **photo-dissociation rate**. This implies that, in the absence of other processes, $[A]$ would decrease exponentially like $\exp(-j_A t)$.

The photo-dissociation rate j_A depends upon the incident flux of solar photons, the frequency ν of the photons and the properties of the molecule A, as follows. Between frequencies ν and $\nu + d\nu$ there is a contribution

$$dj_{Av} = \Phi_{Av}\sigma_{Av} \left(\frac{F_\nu^\downarrow}{h\nu} \right) d\nu,$$

where Φ_{Av} is the **quantum yield** (i.e. the number of reactant molecules decomposed for each absorbed photon) and σ_{Av} is the absorption cross-section for the molecule A

(see [Section 3.5.3](#)). The quantity $F_v^\downarrow / (hv)$ is the incident flux of solar photons, i.e. the number of photons in the given frequency interval crossing unit horizontal area per unit time; it is equal to the spectral irradiance divided by the energy per photon; see [Section 3.6](#). We here assume that the Sun is overhead and that there is no scattering, so that all incident solar photons come from above. For cases of interest to us, the quantum yield is typically quite close to a step function, being zero for frequencies less than a limit ν_0 that corresponds to the minimum energy required to dissociate the molecule and approximately unity for higher frequencies. As shown, for example, in [Figure 3.15](#), the absorption cross-section can vary rapidly with frequency. The spectral irradiance F_v^\downarrow depends on the absorbing gases present in the path of the solar beam; see [Section 3.6](#). The total photo-dissociation rate is found by integration over frequency:

$$j_A = \int_{\nu_0}^{\infty} \frac{\Phi_{Av} \sigma_{Av} F_v^\downarrow}{hv} dv.$$

There is a close relation between the processes of photo-dissociation and thermalisation; see [Section 3.1](#). For a given frequency interval dv , the quantum yield Φ_{Av} equals the rate of consumption of photo-dissociation energy per unit volume, $dj_{Av} [A]hv$, divided by the heating rate per unit volume, $\rho Q_{Av} dv$, say, due to absorption of photons by molecules A. This implies that photo-dissociation rates, like heating rates, may form Chapman layers in the vertical; see [Section 3.6](#).

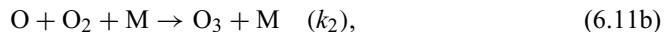
6.5 Stratospheric ozone

6.5.1 Chapman chemistry

The first attempt to explain the presence of the ‘ozone layer’ – that is, the region of maximum ozone number density in the lower stratosphere (see [Section 1.4.1](#)) – was made by [Chapman \(1930\)](#). He used a set of oxygen-only reactions, starting with the photolysis of molecular oxygen by ultra-violet photons of wavelength less than about 240 nm:



cf. reaction [\(6.4\)](#). Next, two fast reactions interconvert O and O₃: the termolecular reaction



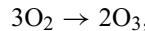
where M is an arbitrary air molecule (cf. reaction [\(6.3\)](#)), and the photolysis of ozone by photons of wavelength less than 1180 nm,



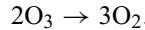
These two reactions determine the partitioning between O and O₃ within **odd oxygen**, defined as O_x = O + O₃. Finally, ozone is destroyed by the slow reaction



Note that the net effect of reactions (6.11a) and (6.11b) is to produce ozone from molecular oxygen,



whereas the net effect of reactions (6.11c) and (6.11d) is to destroy ozone in the reverse manner,



The methods introduced in Sections 6.2 and 6.4 may be applied to reactions (6.11) to determine the equilibrium ozone distribution. On collecting the production and loss terms for ozone, we get

$$\frac{\partial[\text{O}_3]}{\partial t} = k_2[\text{O}][\text{O}_2][\text{M}] - j_3[\text{O}_3] - k_3[\text{O}][\text{O}_3], \quad (6.12\text{a})$$

whereas for atomic oxygen we get

$$\frac{\partial[\text{O}]}{\partial t} = 2j_2[\text{O}_2] + j_3[\text{O}_3] - k_2[\text{O}][\text{O}_2][\text{M}] - k_3[\text{O}][\text{O}_3]. \quad (6.12\text{b})$$

A similar equation can be written down for $\partial[\text{O}_2]/\partial t$, but it is simpler to note that, since the total number of oxygen atoms is conserved by these reactions,

$$[\text{O}] + 2[\text{O}_2] + 3[\text{O}_3] = \text{constant}. \quad (6.12\text{c})$$

Note also that equations (6.12a) and (6.12b) can be added to give an equation for the rate of change of odd oxygen, O_x :

$$\frac{\partial[\text{O}_x]}{\partial t} = \frac{\partial[\text{O}]}{\partial t} + \frac{\partial[\text{O}_3]}{\partial t} = 2j_2[\text{O}_2] - 2k_3[\text{O}][\text{O}_3]. \quad (6.13)$$

In the steady state we set the time derivatives in equations (6.12a), (6.12b) and (6.13) to zero. These equations are not independent, but from two of them, say (6.12a) and (6.13), we can get

$$k_2[\text{O}][\text{O}_2][\text{M}] = j_3[\text{O}_3] + k_3[\text{O}][\text{O}_3], \quad (6.14\text{a})$$

$$j_2[\text{O}_2] = k_3[\text{O}][\text{O}_3]. \quad (6.14\text{b})$$

In the stratosphere it turns out that $j_3 \gg k_3[\text{O}]$, so equation (6.14a) gives the equilibrium ratio of the odd-oxygen species in terms of the rate coefficients j_3 and k_2 and the number densities of O_2 and air (molecule M):

$$\frac{[\text{O}]}{[\text{O}_3]} = \frac{j_3}{k_2[\text{O}_2][\text{M}]} \cdot \quad (6.15)$$

This ratio is determined by the fast reactions (6.11b) and (6.11c); it is found to be very small in the lower and middle stratosphere, so atomic oxygen is much less abundant than is ozone there. (The opposite is true in the upper mesosphere.)

Substitution of equation (6.15) into equation (6.14b) then gives the equilibrium number density of ozone in terms of the number densities of molecular oxygen and air and the rate coefficients:

$$[\text{O}_3] = [\text{O}_2] \left(\frac{j_2 k_2 [\text{M}]}{j_3 k_3} \right)^{1/2}. \quad (6.16)$$

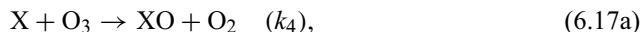
A calculation of the vertical profile of the equilibrium ozone number density predicted by the Chapman theory can therefore be performed if the vertical profiles of the terms on the right-hand side of equation (6.16) are known. It is, however, found that the Chapman theory predicts an ozone maximum that is too large compared with observations (see Problem 6.5). Deficiencies in the Chapman theory include the neglect of further ozone-destroying reactions, particularly catalytic cycles, and the neglect of transport of ozone and other chemicals by atmospheric motions. These are discussed in Sections 6.5.2 and 6.6, respectively.

It is also important to consider the time scales on which equilibrium in the various Chapman reactions is attained. The full details are quite complicated, but we note here that, at 30 km altitude, in the mid-stratosphere, the time scale for equilibration of the odd oxygen ratio (6.15) is a few minutes, whereas the time scale for destruction of odd oxygen is several weeks. There is thus rapid adjustment between O and O₃, but only very slow changes in O + O₃.

6.5.2 Catalytic cycles

As noted in the previous section, the Chapman theory predicts too much ozone in the lower stratosphere. A major reason for this is that it neglects the effects of ozone-destroying **catalytic cycles**. These are now considered briefly.

A common catalytic cycle involves the following pair of reactions:



The net effect of these two reactions is



which destroys ozone just like the Chapman reaction (6.11d). The molecule X is a catalyst; that is, it takes part in the reactions (6.17) but is not itself consumed. Instead it is ‘recycled’, so a single X molecule can destroy many ozone molecules and even tiny amounts of the catalyst may have important effects. Now it was noted in Section 6.5.1 that reaction (6.11d) is slow; if, however, the net effect (6.18) of the catalytic cycle is faster, this will dominate the destruction of ozone, and significantly lower the equilibrium amount (see Problem 6.6). Examples of catalysts X that actually do this in the stratosphere are the hydroxyl radical (OH), nitric oxide (NO) and chlorine (Cl).

Inclusion of catalytic cycles brings theoretical calculations of the ozone distribution into closer agreement with observations. However, the cycles interact with one another, so their effects cannot be added linearly. Further complications include the facts that some of the catalysts may become involved in additional ‘null cycles’, making them unavailable to reactions (6.17). Moreover, some potential catalysts may be temporarily tied up in unreactive ‘reservoir species’, such as dinitrogen pentoxide (N_2O_5), hydrochloric acid (HCl) and chlorine nitrate (ClONO_2). Comprehensive models of stratospheric chemistry must take all of these effects into account and hence may include tens of reactants and hundreds of reactions.

6.6 The transport of chemicals

It was noted above that a full theory of stratospheric ozone needs to take account of the transport of chemicals by air motion; in particular, we must apply the equations governing chemical reactions to moving parcels of air. An important concept here is the material, or advective, derivative D/Dt , introduced in Section 4.2, which represents the rate of change with time following a moving parcel.

Consider an air parcel of unit mass, including a chemical constituent A whose mass mixing ratio is μ_A . Then by definition of the mass mixing ratio (see Section 2.2), μ_A is also the mass of A in the parcel. If A is neither being produced nor being lost through chemical reactions, the mass of A is constant following the parcel, so

$$\frac{D(\mu_A)}{Dt} = 0. \quad (6.19)$$

A chemical whose mixing ratio satisfies equation (6.19) is an example of a **conservative tracer**.²

It is straightforward to show that the mass mixing ratio of A is related to the number density [A] by

$$\rho\mu_A = m_A[A], \quad (6.20)$$

where ρ is the density of air, including the constituent A (see Section 2.2), and m_A is the molecular mass of A. Equation (6.19) can therefore be written

$$\frac{D}{Dt}\left(\frac{m_A[A]}{\rho}\right) = 0$$

and so, since m_A is constant,

$$\frac{1}{\rho} \frac{D[A]}{Dt} - \frac{[A]}{\rho^2} \frac{D\rho}{Dt} = 0.$$

² Other examples of conservative tracers are the potential temperature θ (when the diabatic heating $Q = 0$; see Section 4.10) and the potential vorticity P (when both the diabatic heating and friction are zero; see Section 5.1).

However, from the mass-continuity equation (4.8), $D\rho/Dt = -\rho\nabla \cdot \mathbf{u}$, where \mathbf{u} is the fluid velocity. After some rearrangement and expansion of the material derivative $D/Dt = \partial/\partial t + \mathbf{u} \cdot \nabla$, we get

$$\frac{\partial[A]}{\partial t} + \mathbf{u} \cdot \nabla[A] + [A]\nabla \cdot \mathbf{u} = 0,$$

or, collecting the terms in \mathbf{u} ,

$$\frac{\partial[A]}{\partial t} + \nabla \cdot (\mathbf{u}[A]) = 0. \quad (6.21)$$

This shows how the number density of a conservative tracer A changes in response to compression or expansion of a parcel of air, as it is advected with the local velocity field \mathbf{u} . The analogy with the mass-continuity equation (4.3) should be noted; see Problem 6.8.

Now suppose that A is undergoing chemical reactions so that, in addition to this compression or expansion effect, there is a contribution $-R_A$ to the rate of change of the number density [A]; cf. equation (6.6). Equation (6.21) must then be replaced by

$$\frac{\partial[A]}{\partial t} + \nabla \cdot (\mathbf{u}[A]) = -R_A. \quad (6.22)$$

All time derivatives of number densities in Sections 6.2–6.5 must be replaced by expressions analogous to that on the left-hand side of equation (6.22), when transport is to be taken into account. The right-hand side of equation (6.22) is commonly split into ‘production’ and ‘loss’ terms in the form $-R_A = P_A - L_A$.

It can be shown that, when chemical reactions are taking place, equation (6.19) for the mixing ratio must be replaced by

$$\frac{D(\mu_A)}{Dt} = -\frac{m_A R_A}{\rho}. \quad (6.23)$$

In this case A is called a **non-conservative tracer**.

By analogy with the chemical lifetimes defined in Section 6.2, we can also define a **dynamical lifetime**

$$\tau_{\text{dyn}} = \frac{[A]}{|\nabla \cdot (\mathbf{u}[A])|}.$$

This may be difficult to estimate without quite detailed knowledge of the velocity field, but, in general, if τ_{dyn} is much longer than the chemical lifetime τ_A , then changes in A will be mainly chemical, whereas if τ_{dyn} is much shorter than τ_A then A will be essentially a conservative tracer.

In principle, a full treatment of atmospheric chemistry should follow the motion of each parcel of air and account for the changes of each chemical constituent within the parcel, using equation (6.22) or equation (6.23). In practice this would be impossibly complicated, even if the motion itself were known precisely. One feasible approach is to follow the trajectories of a small but representative set of parcels and track the chemical changes for each parcel as it moves through regions of differing composition and perhaps into and out of sunlight. Such a trajectory calculation needs accurate knowledge of the wind fields and

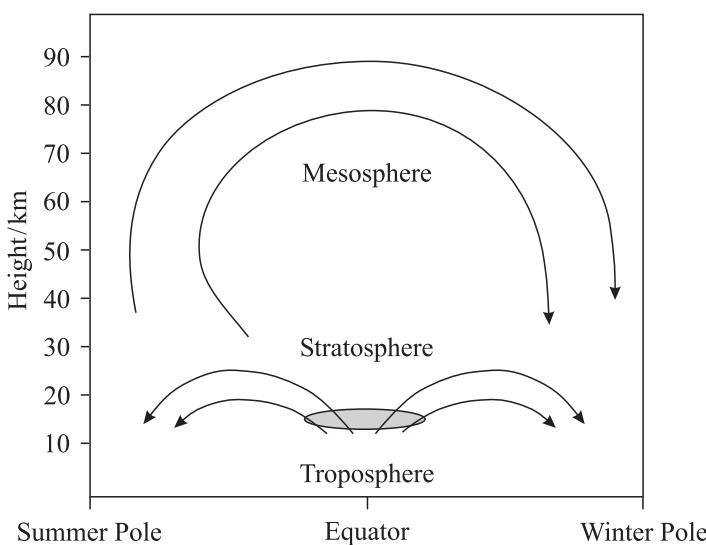


Fig. 6.4

A schematic plot of the Brewer–Dobson circulation (lower four arrows, in the stratosphere) and the solstitial mesospheric circulation (upper two arrows). The shaded ellipse indicates the approximate position of the 200 K isotherm near the equatorial tropopause, i.e. the ‘cold trap’ identified by [Brewer \(1949\)](#).

these must usually be found from models: observations on their own are seldom sufficient for an accurate determination of winds over extensive regions.

Nevertheless, some patterns of transport are well established, at least in a qualitative sense. The most important transport structure in the stratosphere is the **Brewer–Dobson circulation**: this is a meridional circulation, i.e. a circulation in the north–south and vertical directions. (The superimposed east–west motion is much more rapid, but is conceptually less important for global transport of chemicals.) It involves upward motion of air parcels from the troposphere into the stratosphere at low latitudes, poleward motion in the lower stratosphere of each hemisphere and a gentle descent back into the troposphere in middle and high latitudes; see [Figure 6.4](#). The upward motion in low latitudes was inferred by [Brewer \(1949\)](#) from the fact that the stratosphere is very dry: he reasoned that the inability of moisture from the troposphere to get into the stratosphere must be due to the fact that most of the air that moves from the troposphere into the stratosphere passes through a very cold region (often called the ‘cold trap’ in this context) at the equatorial tropopause (see [Figure 1.5](#)). Here it is ‘freeze-dried’, with water vapour forming ice particles that fall out under gravity. The poleward motion in the lower stratosphere was inferred by [Dobson \(1956\)](#) from observations of the ozone distribution: although ozone is mainly formed in the equatorial stratosphere from reactions (6.11a) and (6.11b), significant number densities are found at high latitudes, which must be due to transport from low latitudes. (Note that, if a parcel retains its ozone mixing ratio as it descends – and is compressed – at high latitudes, then the number density increases. This is also evident from [equation \(6.20\)](#), since the parcel encounters increasing air density as it descends.) Although the general form of the Brewer–Dobson circulation was originally obtained from the observed distributions of

chemicals, elucidation and interpretation of its structure has required detailed theoretical and numerical modelling studies.

Another large-scale circulation in the middle atmosphere occurs around the solstices, with upward motion in the summer stratosphere and mesosphere, a summer-to-winter motion in the mesosphere and a descent over the winter pole; see [Figure 6.4](#) again. This circulation also has important implications for the transport of chemicals. Both of these circulations are essentially wave-driven: the Brewer–Dobson circulation by planetary waves and the solstitial circulation mainly by gravity waves (see also [Section 9.5.2](#)).

6.7 The Antarctic ozone hole

A useful general measure of ozone in the atmosphere is the **column ozone** or **total ozone**, defined in terms of the total number N_3 of molecules of ozone in a complete vertical column of atmosphere of unit horizontal cross-section:

$$N_3 = \int_0^\infty [\text{O}_3] dz.$$

A convenient measure of column ozone is the **Dobson Unit** (DU), defined as the height of the column, in hundredths of a millimetre, if all the ozone molecules in it were brought to a pressure of 1 atm and a temperature of 0 °C. Typical values of column ozone are about 300 DU, meaning that the height of the column, compressed in this way, would be only about 3 mm. Most of the ozone molecules occur in the stratosphere.

As was mentioned in [Section 1.4.4](#) and illustrated in [Figure 1.10](#), the column ozone varies from place to place over the globe and with time; we focus here on its behaviour in the Antarctic. Measurements of the seasonal variation of column ozone amounts over Antarctica have been made since 1956. In particular, members of the British Antarctic Survey have made such measurements with a Dobson ozone spectrophotometer (see [Section 7.3.1](#)) at Halley station (76° S, 27° W), except during the midwinter polar night, when the Sun does not rise for almost 4 months. The typical annual variation in the Antarctic, prior to the late 1970s, had a minimum of about 250–300 DU in spring and a maximum of about 400 DU in summer. However, in more recent years the picture has changed significantly: monthly mean October amounts generally decreased throughout the 1980s and have fallen below 160 DU in each year between 1991 and 2008, except 2002 and 2004; see [Figure 6.5](#). These routine ground-based measurements were confirmed by satellite observations with the Total Ozone Mapping Spectrometer (TOMS) and also by several special international expeditions to the Antarctic, in which very detailed measurements of the spring ozone-depletion phenomenon were made. These revealed a large-scale ‘ozone hole’, a region whose horizontal size approaches that of the Antarctic continent, within which a massive loss of ozone occurs over a period of about 6 weeks during the Southern Hemisphere’s spring, between August and October. Observations of the vertical structure of the ozone layer in the Antarctic show that, at many locations, nearly all the ozone may disappear at this time between about 15 and 20 km altitude; see [Figure 6.6](#). Later in the year, ozone

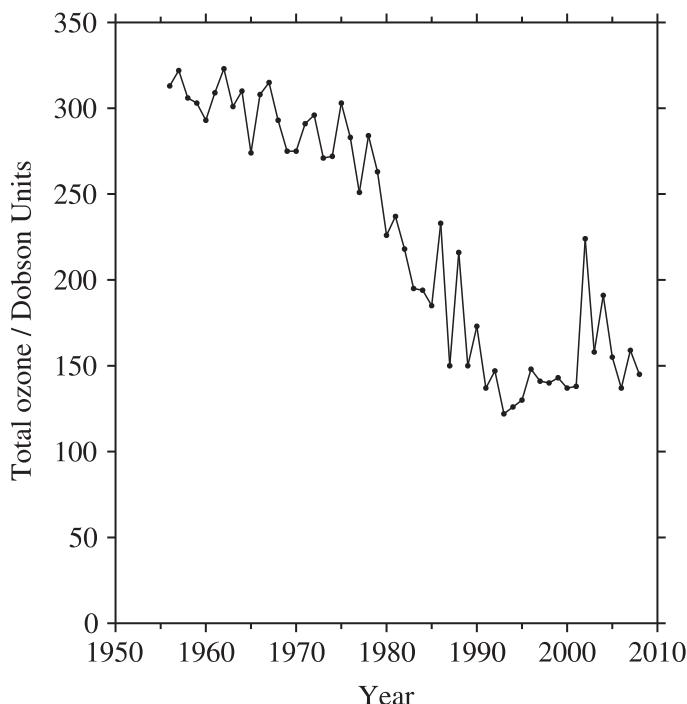


Fig. 6.5 Monthly mean values of column ozone in October at Halley station for 1957–2008. Drawn using data provided by J. D. Shanklin, British Antarctic Survey.

returns to the Antarctic stratosphere, although the region of depleted ozone content may drift away into lower latitudes before it is ‘filled in’.

After extensive study by many atmospheric chemists and physicists a broad understanding of the Antarctic ozone hole has been attained: a qualitative description is as follows. In the winter night the polar regions receive no solar radiation, but can readily emit infrared radiation to space. At the same time a strong circumpolar (eastward) vortex forms in the Antarctic stratosphere; this is subject to very little disturbance by planetary waves³ and so there is little wave-induced dynamical heat transport from lower latitudes into the vortex. The vortex therefore becomes very cold and ‘polar stratospheric clouds’ (PSCs) – thin clouds of ice or liquid water particles – can form in the Antarctic lower stratosphere. Similarly, there is little latitudinal transport of chemicals into or out of the vortex, which therefore acts rather like a ‘chemical containment vessel’. By contrast, the Arctic polar winter vortex is usually more disturbed by planetary waves and has more heat transport

³ The Northern Hemisphere has two large continental land-masses at midlatitudes: during the winter the eastward winds blowing over these continents give rise to stationary Rossby waves in the troposphere (see Section 5.5). These in turn lead to stratospheric Rossby-wave-like disturbances that disrupt the northern winter stratospheric vortex from time to time. The absence of large midlatitude continental land-masses in the Southern Hemisphere means that such disruptions seldom occur there.

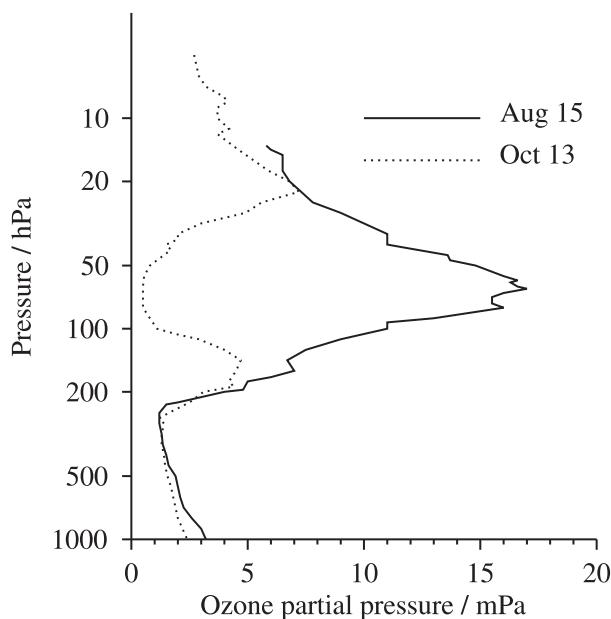


Fig. 6.6 The partial pressure of ozone in the lower stratosphere over Halley station on 15 August (solid line) and 13 October 1987 (dotted), showing the almost total loss of ozone between about 15 and 20 km altitude during this period. Adapted from [Gardiner \(1989\)](#).

from lower latitudes, so the cold winter conditions that allow PSCs to form are less common in the Arctic stratosphere.

It turns out that purely gas-phase chemistry cannot account for the large ozone losses that are found in the lower stratosphere during the Antarctic spring: ‘heterogeneous’ reactions, on the surfaces of the PSC particles, are also necessary. (The vertical region of ozone depletion shown in Figure 6.6 is located near a PSC.) The process is too complex to be explained in detail here, but the main features are as follows.

- Catalytic cycles involving chlorine species, including the chlorine monoxide dimer (ClO_2), and reactive bromine species (e.g. BrO), can effectively remove ozone at the cold temperatures present in the lower Antarctic stratosphere during the spring. These cycles are inhibited if ClO becomes trapped in the reservoir species chlorine nitrate (ClONO_2), as noted in Section 6.5.2. However, this effect is diminished in the presence of PSCs, since nitric acid (HNO_3) condenses on the surfaces of the PSC particles, which in turn may sediment out under gravity, removing the NO_2 that is needed to form ClONO_2 .
- Moreover, the reservoir species ClONO_2 and HCl (hydrochloric acid) can actually release chlorine compounds (Cl_2 and HOCl) on the surface of the PSCs and these compounds are readily photolysed when the Sun returns in early spring, providing a supply of Cl for the catalytic cycles.
- A chlorine catalytic cycle of the form (6.17), where $\text{X} = \text{Cl}$, is too slow in the lower stratosphere to account for the observed depletion of ozone.

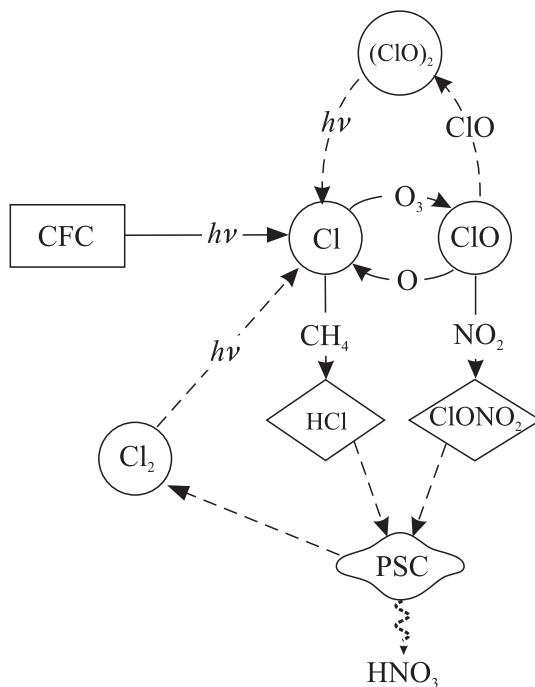
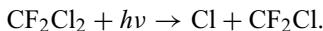


Fig. 6.7 A schematic diagram of some of the chemical processes involved in ozone depletion in the Antarctic. The rectangle represents the source of chlorine, the circles represent reaction intermediates and the diamonds represent reservoir species. The dashed lines represent reactions that occur only in the presence of PSCs or at unusually low temperatures. Adapted from Wayne (2000).

- Some of the stratospheric chlorine is derived from naturally occurring methyl chloride (CH_3Cl), but most comes from photolysis in the stratosphere of the man-made chlorofluorocarbons (CFCs); for example, the CFC CF_2Cl_2 is photolysed by ultra-violet radiation in the reaction



A schematic diagram of these processes is given in Figure 6.7.

The Antarctic ozone hole therefore arises from a blend of chemical, dynamical and transport processes. Transport carries chlorine compounds from the industrial regions of the Northern Hemisphere to the Antarctic. The absence of planetary waves makes the Antarctic winter stratosphere an especially cold and isolated part of the atmosphere, where PSCs can form. Finally, chemical reactions, including reactions on the PSC particles, lead to rapid and massive local depletion of ozone in the Antarctic spring.

Ozone depletion is not confined to the Antarctic spring. World-wide data show that there have been ozone losses at other latitudes since the 1970s: for example a 5% per decade drop in column ozone has been observed between 40°N and 60°N in winter. Possible causes of global-scale depletion include the ‘dilution’ of the rest of the atmosphere, due

to dispersion of the Antarctic ozone hole when the vortex finally breaks down in spring; similar but weaker ‘mini ozone holes’ in the Arctic, which have been observed in some recent winters; and other types of heterogeneous chemistry, taking place on volcanic cloud particles and involving further chemical species.

The decrease in amount of atmospheric ozone may also have climatic effects, in addition to the direct biological effects mentioned in [Section 1.4.4](#). Less ozone in the stratosphere implies less absorption of solar and infra-red radiation there (see [Section 3.5.3](#)) but also less infra-red emission: the net effect is found to give a cooler stratosphere. Moreover, the decrease in amount of stratospheric ozone during recent decades is believed to have led to cooling in the troposphere also, partly offsetting the enhancement of the greenhouse warming there due to the increase in amounts of carbon dioxide and other greenhouse gases: see [Section 8.1](#).

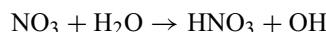
Further reading

The book by [Graedel and Crutzen \(1995\)](#) provides a non-technical introduction to atmospheric chemistry and its implications for climate. The basic physical chemistry mentioned in this chapter is treated in more detail by [Atkins \(2006\)](#) and [Hobbs \(1995\)](#), for example, and the thermodynamic and kinetic-theory background are covered, for example, by [Blundell and Blundell \(2009\)](#). Specialised texts on atmospheric chemistry, covering most of the topics in this chapter, include those by [Wayne \(2000\)](#) and [Brasseur and Solomon \(2005\)](#). [Lary \(1997\)](#) presents a review of ozone-destroying catalytic cycles in the stratosphere. A detailed treatment of stratospheric transport is given by [Andrews *et al.* \(1987\)](#). [Dessler \(2000\)](#) provides an introductory survey of the chemistry and physics of stratospheric ozone while [Solomon \(1999\)](#) gives an authoritative review of research into stratospheric ozone depletion.

Problems

Problem 6.1 Verify the result, mentioned in [Section 6.1](#), that the photolytic reaction (6.4) requires photons of wavelength less than about 240 nm, given that the molar enthalpy of reaction is about $+500 \text{ kJ mol}^{-1}$.

Problem 6.2 Can the reaction



occur spontaneously at standard temperature and pressure? Use the following Gibbs free energies of formation at STP (all in kJ mol^{-1}): HNO_3 , -74.7 ; OH , 34.2 ; NO_3 , 115.8 ; and H_2O , -228.4 .

Problem 6.3 Consider a bimolecular reaction in the form of [equation \(6.1\)](#). Show from [equation \(6.7\)](#) that $[\text{B}] - [\text{B}]_0 = [\text{A}] - [\text{A}]_0$, where the subscripts 0 indicate the values at

$t = 0$. Then solve for $[A]$, assuming that the rate coefficient k_{AB} is constant. Hence verify that [equation \(6.8\)](#) holds if $[B]_0 \gg [A]_0$.

Problem 6.4 If $F_s = 1370 \text{ W m}^{-2}$ is the total solar irradiance and P_s is the total solar photon flux ($\text{photons m}^{-2} \text{ s}^{-1}$), both measured at the top of the atmosphere, show that

$$\frac{P_s}{F_s} = \frac{\int_0^\infty N(\nu) d\nu}{\int_0^\infty B_\nu(\nu, T) d\nu},$$

where $N(\nu) d\nu$ is the number of photons $\text{m}^{-2} \text{ s}^{-1}$ steradian $^{-1}$ emitted from the Sun's surface in the frequency range $(\nu, \nu + d\nu)$, and $B_\nu(\nu, T)$ is the Planck function; assume that the Sun is a black body at $T = 6000 \text{ K}$. Hence obtain P_s , given that

$$\int_0^\infty \frac{x^2}{e^x - 1} dx = 2.40.$$

Calculate also the fraction of the total photon flux that lies in each of the wavelength ranges (i) $\lambda < 1180 \text{ nm}$, (ii) $240 \text{ nm} < \lambda < 280 \text{ nm}$ and (iii) $\lambda < 240 \text{ nm}$. (You will need to write a computer program to calculate the required integrals.)

Problem 6.5 The following reaction rates and photo-dissociation rates are taken from [DeMore et al. \(1997\)](#):

$$k_2 = 6 \times 10^{-46} (T/300)^{-2.3} \text{ m}^6 \text{ s}^{-1},$$

$$k_3 = 8 \times 10^{-18} \exp(-2060/T) \text{ m}^3 \text{ s}^{-1},$$

$$j_2 = 3 \times 10^{-12} \text{ s}^{-1}, \quad j_3 = 5.5 \times 10^{-4} \text{ s}^{-1}.$$

The photo-dissociation rates are appropriate for the midlatitude equinox at noon and at 25 km altitude. Use information from [Figure 1.3](#) and [Section 2.2](#) to estimate the number density of ozone at this altitude, based on Chapman chemistry. Show that this value is several times larger than the peak value given in [Figure 1.4](#). What physical and chemical processes might reduce this value in the real atmosphere?

Problem 6.6 Add the catalytic reactions [\(6.17\)](#) to the Chapman scheme and explain how this modifies the algebra for the Chapman scheme. Assuming that the extra reactions are much slower than the fast equilibrium between O_3 and $\text{O}_2 + \text{O}$, show that the vertical distribution of ozone is given by

$$\frac{[\text{O}_3]}{[\text{O}_2]} \approx \left(\frac{j_2 k_2 [\text{M}]}{j_3 k_3} \right)^{1/2} - \left(\frac{k_2 k_4 [\text{M}][\text{X}]}{2 j_3 k_3} \right);$$

compare this with [equation \(6.16\)](#) and comment.

Problem 6.7 What happens to O and O_3 , according to the Chapman scheme, when solar radiation switches off at sunset? If catalysts are present, what happens to X and XO ? (Assume that $k_2[\text{O}_2][\text{M}] + k_3[\text{O}_3] \geq 10^{-1} \text{ s}^{-1}$ throughout the stratosphere.)

Problem 6.8 Derive equation (6.21) by a direct method analogous to that used in Chapter 4 to derive the continuity equation (4.3).

Problem 6.9 Check that 1 DU corresponds to an ozone column of about 2.7×10^{20} molecules per square metre.

In this chapter we consider a small selection of techniques for observing the atmosphere. These techniques have been chosen for two main reasons: (a) they illustrate the use of physical principles, including principles introduced earlier in this book; and (b) they provide crucial data on atmospheric phenomena modelled elsewhere in this book, such as Rossby waves, gravity waves and the Antarctic ozone hole. The topics considered are all examples of **remote sounding**; we do not attempt to present a balanced account of all observational methods.

In Section 7.1 we briefly list some of the main atmospheric observational methods. In Section 7.2 we outline the principles of remote sounding of the atmosphere from space, focusing on methods that rely on thermal emission from atmospheric gases and on scattering of solar radiation by atmospheric gases. Then in Section 7.3 we discuss three types of ground-based remote sounding, namely the Dobson spectrophotometer, radars and lidars. We omit the details of the instruments' optical and electronic systems, the technicalities of signal processing and the sophisticated statistical methods that may be required in order to extract meaningful physical quantities from the raw measurements.

7.1 Atmospheric observations

Quantitative observations of the atmosphere are made in many different ways. Routine meteorological measurements of ground-level temperature and wind are made with simple thermometers and anemometers, respectively, and routine measurements of temperature and humidity through the depth of the troposphere are made with balloon-borne instruments (radiosondes) that transmit information back to the surface by radio. At higher altitudes, instruments on research balloons and aircraft can measure samples of air in the lower stratosphere, and instruments on rockets can make measurements in the mesosphere and beyond. These are all examples of *in situ* measurements. An alternative is to use **remote sounding**, that is, to measure atmospheric properties at a distance, rather than in the vicinity of the measuring apparatus. This can be done from orbiting instruments in space, from aircraft, from balloons and from ground-based instruments, in each case by measuring electromagnetic radiation emitted, scattered or transmitted by the atmosphere.

Remote-sounding techniques can be divided into **passive** and **active** types. In passive remote sounding, the radiation measured is of natural origin, for example thermal radiation emitted by the atmosphere, or solar radiation transmitted or scattered by the atmosphere. (Most space-borne remote-sounding methods are passive.) In active remote sounding, a

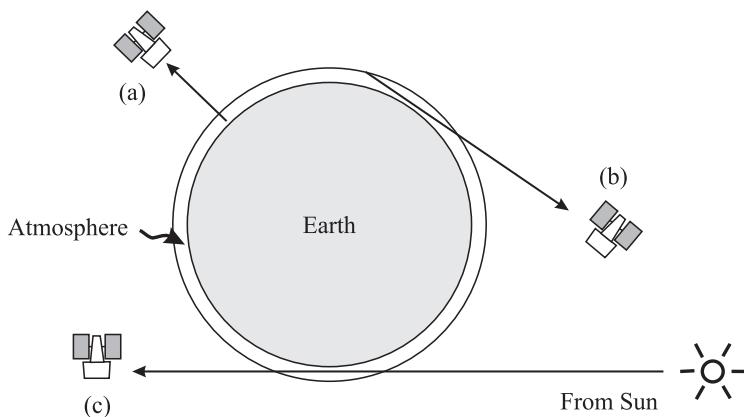
transmitter (e.g. a radar) is used to direct pulses of radiation into the atmosphere, where they are scattered by atmospheric molecules, aerosols or inhomogeneities of atmospheric structure; some of the scattered radiation is then detected by a receiver.

Each of these techniques has its advantages and disadvantages. *In situ* measurements may give accurate, high-resolution measurements, but can sample only small, perhaps unrepresentative, regions. Remote sounding from satellites can give near-global coverage, but can provide only averaged values of the measured quantity over large regions, perhaps hundreds of kilometres in horizontal extent and several kilometres in the vertical direction. Satellite instruments are expensive to put into orbit and cannot usually be repaired if they fail. Ground-based radars can provide data with high vertical resolution (by measuring small differences in the time delays of the return pulses), but only above the radar site. In general a mix of different observational techniques must be used for providing the data required for weather forecasting or atmospheric research.

7.2 Atmospheric remote sounding from space

Earth-orbiting satellites provide enormous possibilities for global-scale measurements of the atmosphere. For example, geostationary satellites, orbiting at 36 000 km over the equator, monitor weather systems and provide information on tropospheric winds by tracking the motion of clouds. Polar-orbiting and other non-geostationary satellites, orbiting at altitudes up to about 1000 km, are used for measuring temperature and composition at a range of levels in the vertical. They complete up to 15 orbits per day, crossing each latitude circle twice per orbit; their orbits are often chosen such that they remain ‘phase-locked’ with the Sun (they are then called **Sun-synchronous**), crossing each latitude at the same two local times in each orbit.

In the following subsections we shall consider some basic principles of temperature and composition sounding from space, focusing on measurements of the thermal radiation emitted from the atmosphere and of solar radiation scattered by the atmosphere. Two viewing geometries are commonly used (see [Figure 7.1](#)): **nadir viewing**, in which the instrument views vertically downward towards the ground, and **limb viewing**, in which the instrument views the atmosphere tangentially, towards the limb. Nadir viewing measures radiation from a comparatively short path-length of emitting gas, against the warm background of the Earth’s surface and lower atmosphere, which also emit thermal radiation. Limb viewing has the advantage of measuring radiation from a much longer path-length of emitting gas (see [Problem 7.1](#)), against the non-emitting background of cold space. However, the spacecraft must be aligned very accurately (to within a few arc seconds) for limb viewing, if the correct portion of atmosphere is to be measured. The technique of solar occultation, in which measurements of the absorption of solar radiation by the atmosphere are made as the Sun rises or sets, is also indicated in [Figure 7.1](#), but will not be considered in this chapter. All of these techniques make use of the ideas of radiative transfer introduced in [Chapter 3](#).

**Fig. 7.1**

Illustrating the geometry of (a) nadir viewing, (b) limb viewing and (c) solar occultation. (Not to scale.) Note that the limb-viewing and occultation instruments may observe a different latitude from the current latitude of the satellite platform; for example, the North Pole may be viewed while the satellite is at 60° N.

7.2.1 Thermal emission measurements

Consider a satellite instrument receiving infra-red emission from the atmosphere along a vertical (nadir) path. From the solution of the radiative-transfer equation in the form of equation (3.13), assuming local thermodynamic equilibrium so that $J_\nu = B_\nu$, we can show that the spectral radiance L_ν at frequency ν received by the satellite is given by

$$L_\nu = \int_0^\infty B_\nu(T(z)) \frac{\partial T_\nu(z, \infty)}{\partial z} dz + B_\nu(T_g) T_\nu(0, \infty), \quad (7.1)$$

where $T(z)$ is the atmospheric temperature at height z , $T_\nu(z, \infty)$ is the spectral transmittance between height z and the satellite, and T_g is the surface temperature. It is assumed that the surface acts as a black body.¹

The spectral absorptance between height z and the satellite is $A_\nu(z, \infty) = 1 - T_\nu(z, \infty)$: for simplicity the ‘ ∞ ’ argument of A_ν will be dropped from now on. The general form of A_ν near an isolated spectral line at $\nu = \nu_0$ is given by Figure 7.2 (cf. Figure 3.12). The precise details depend on whether we have a Doppler or Lorentz line shape, on the amount and distribution of absorbers or scatterers in the path between height z and the satellite, and on the temperature and pressure distributions along the path.

The form of the ν dependence of the Planck function $B_\nu(T)$, a much more slowly varying function of ν than the absorptance, is given in Figure 7.3. Note that B_ν increases with T for all values of ν and T .

The simplest case of remote sounding of temperature occurs when atmospheric extinction at frequency ν can be neglected. Then $T_\nu = 1$ everywhere and the satellite just ‘sees the surface’ at that frequency, with no atmospheric extinction, so $L_\nu = B_\nu(T_g)$. The surface

¹ Note the resemblance of equation (7.1) to equation (3.33); however, the latter involves integration over all slanting upward paths.

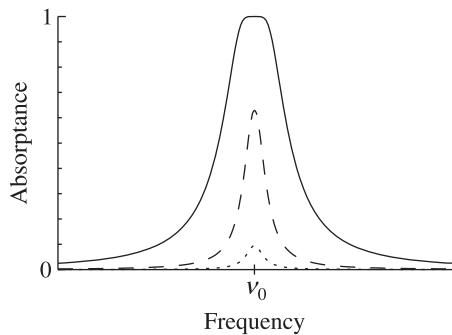


Fig. 7.2 Schematic form of the spectral absorptance $\mathcal{A}_v(z)$ between height z and a remote-sounding satellite, as a function of frequency v near a Lorentz line centre at v_0 , for three values of z . Solid curve, small z ; dashed curve, intermediate z ; dotted curve, large z .

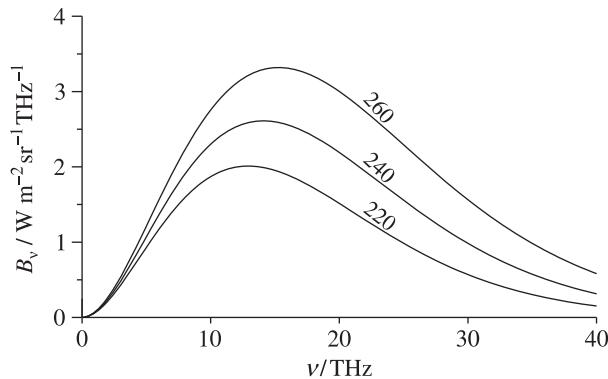


Fig. 7.3 Form of the Planck function $B_v(T)$, as a function of frequency v , for the temperatures $T = 220, 240$ and 260 K . The peak of the 260 K curve, at about 16 THz , corresponds to a wavenumber of about 530 cm^{-1} and a wavelength of about $18.7\text{ }\mu\text{m}$. Note that $1\text{ THz} = 10^{12}\text{ Hz}$.

temperature T_g can then be obtained from the radiance measurement L_v at frequency v by inverting the Planck function (3.1):

$$T_g = \frac{hv}{k_B \ln \left(1 + \frac{2hv^3}{c^2 L_v} \right)}. \quad (7.2)$$

More generally, we can define a **brightness temperature** in terms of the radiance L_v and frequency v using equation (7.2), even in the presence of extinction.

Now suppose that the frequency v is near an isolated spectral line centred at v_0 , so that absorption cannot be neglected, and consider an isothermal atmosphere at a constant temperature T_0 , which may be different from the temperature T_g of the ground. Then the integral in equation (7.1) can be evaluated explicitly, to give

$$\begin{aligned}
 L_\nu &= B_\nu(T_0)[1 - T_\nu(0, \infty)] + B_\nu(T_g)T_\nu(0, \infty) \\
 &= B_\nu(T_0)\mathcal{A}_\nu(0) + B_\nu(T_g)[1 - \mathcal{A}_\nu(0)] \\
 &= B_\nu(T_g) + [B_\nu(T_0) - B_\nu(T_g)]\mathcal{A}_\nu(0).
 \end{aligned} \tag{7.3}$$

Now $1 - T_\nu(0, \infty) = \mathcal{A}_\nu(0)$ is the spectral absorptance, and hence also the spectral emittance, of the isothermal atmosphere. The first and second lines of equation (7.3) therefore state that L_ν equals the upward radiance $B_\nu(T_0)\mathcal{A}_\nu(0)$ from the (non-black) atmosphere, plus a fraction $T_\nu(0, \infty)$ of the upward radiance $B_\nu(T_g)$ from the (black) ground that is transmitted by the atmosphere. The last line of equation (7.3) shows that the radiance L_ν can alternatively be written as $B_\nu(T_g)$ plus a modification equal to $B_\nu(T_0) - B_\nu(T_g)$ times the total atmospheric absorptance $\mathcal{A}_\nu(0)$. Note that any rapidly varying frequency dependence of L_ν must come from the absorptance \mathcal{A}_ν , rather than from the Planck function B_ν .

Consider two cases.

- If $T_0 > T_g$ (the air temperature is greater than the surface temperature), then $B_\nu(T_0) > B_\nu(T_g)$ and L_ν displays a ‘hump’ around the line centre at ν_0 , as shown in Figure 7.4. The line is called an **emission line**.
- If $T_0 < T_g$ (the surface is warmer than the air temperature), then L_ν displays a ‘dip’ around the line centre at ν_0 , as shown in Figure 7.5. The line is called an **absorption line**.

In each case the radiance L_ν tends to the Planck function at the surface temperature, $B_\nu(T_g)$, sufficiently far from the line centre. The depth of the hump or dip at ν_0 in these two cases increases as the absorptance $\mathcal{A}_\nu(0)$ increases and also as the magnitude of the jump $B_\nu(T_0) - B_\nu(T_g)$ in the Planck function increases.

Now consider the two-layer atmospheric temperature profile given by

$$T(z) = T_0 \quad \text{for } 0 < z < z_1, \quad T(z) = T_1 \quad \text{for } z_1 < z.$$

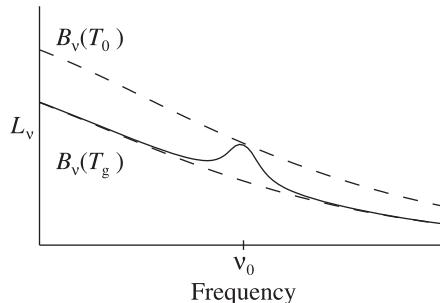


Fig. 7.4

Schematic satellite-observed radiance (solid line) near a Lorentz-broadened spectral line, for an isothermal atmosphere that is warmer than the surface temperature. Planck functions at the temperatures T_g and T_0 are given as dashed lines.

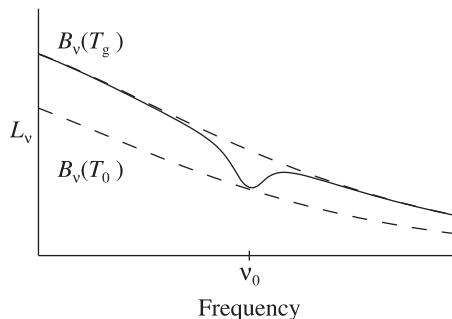


Fig. 7.5

Schematic satellite-observed radiance (solid line) near a Lorentz-broadened spectral line, for an isothermal atmosphere that is colder than the surface temperature. Planck functions at the temperatures T_g and T_0 are given as dashed lines.

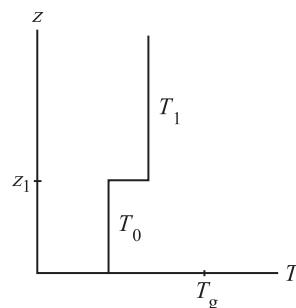


Fig. 7.6

Schematic temperature profile for a two-layer atmosphere.

It is easy to show from equation (7.1) that

$$L_v = B_v(T_g) + [B_v(T_0) - B_v(T_g)] \mathcal{A}_v(0) + [B_v(T_1) - B_v(T_0)] \mathcal{A}_v(z_1).$$

Suppose for example that $T_g > T_1 > T_0$, as in Figure 7.6, crudely representing a hot surface, cool ‘troposphere’ and warm ‘stratosphere’. The frequency variation of the resulting radiance near v_0 received at the satellite takes the form of Figure 7.7, in the case of a Lorentz-broadened spectral line. The spectral radiance L_v includes a ‘background’ term given by the Planck function at the surface temperature, $B_v(T_g)$. This has a large dip superimposed: large, because $\mathcal{A}_v(0)$ is a broad, high hump, and a dip because $B_v(T_0) < B_v(T_g)$. In the middle of this large dip is a small hump: small because $\mathcal{A}_v(z_1)$ is generally lower and narrower than $\mathcal{A}_v(0)$ and also because the jump in B_v at z_1 is less than at the ground, and a hump because $B_v(T_1) > B_v(T_0)$.² This kind of feature can be seen in the atmospheric example shown in Figure 7.8; see also Problem 7.3.

A similar approach can be used for a multi-layer atmosphere, approximating a smoothly varying profile $T(z)$ by a series of steps. If there are many layers, however, it is difficult

² The wideness of the dip and narrowness of the hump result from the temperature and pressure dependences of the Lorentz width γ_L : see equation (3.22).

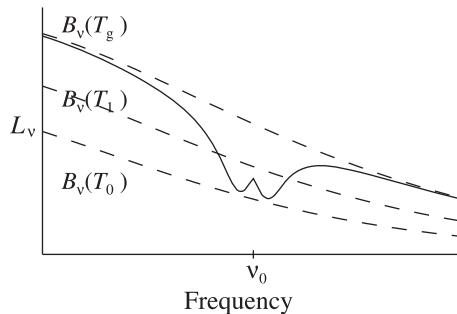


Fig. 7.7

Schematic satellite-observed radiance (solid line) for the two-layer atmosphere shown in Figure 7.6. Planck functions at the temperatures T_g , T_0 and T_1 are given as dashed lines. It is assumed that $\mathcal{A}_v(0) \approx 1$ and $\mathcal{A}_v(z_1) \approx 0.6$ at the central frequency v_0 .

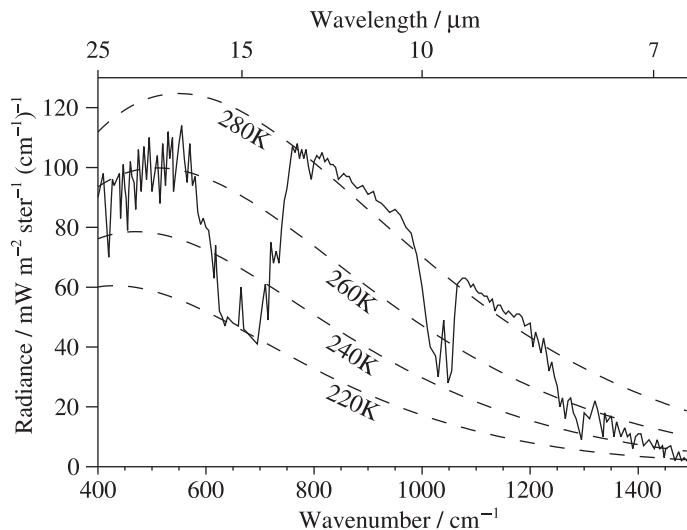


Fig. 7.8

Infrared radiance measurements over the Mediterranean Sea from the IRIS instrument on the Nimbus 4 satellite. Thermal emission from the atmosphere and the surface (solid curve) is shown as a function of wavenumber (in units of cm^{-1}) and wavelength (in units of μm), together with the Planck function at several temperatures (dashed curves). Adapted from Hanel *et al.* (1971).

to draw simple insights from the resulting spectral data; we therefore adopt a different approach, as follows. For convenience, we first introduce the log-pressure coordinate

$$Z = \ln(p_0/p), \quad (7.4)$$

where p_0 is the surface pressure. Recall from equation (2.15) that, for an isothermal atmosphere at temperature T_0 , the pressure decreases exponentially with height, $p = p_0 \exp(-z/H)$, where $H = RT_0/g$ is the pressure scale height ($H \approx 7.6 \text{ km}$ for $T_0 = 260 \text{ K}$). Therefore in this isothermal case $Z = z/H$. Although Z is not precisely proportional to height for a non-isothermal atmosphere, it is still approximately so in the lower and middle atmospheres. The quantity Z is sometimes called the ‘log-pressure in scale heights’.

On changing variables from z to Z in the integral in equation (7.1), we obtain

$$L_\nu = \int_0^\infty B_\nu(T(Z)) \mathcal{K}(Z) dZ + B_\nu(T_g) T_\nu(0, \infty), \quad (7.5)$$

where

$$\mathcal{K}(Z) = \frac{\partial T_\nu(Z, \infty)}{\partial Z} \quad (7.6)$$

is the **weighting function**; it weights the contributions to L_ν from the Planck functions at different log-pressure altitudes Z .

In certain simple cases, the weighting function can be evaluated explicitly. First note from the definition (3.23) of the transmittance that, in terms of height,

$$T_\nu(z, \infty) = \exp\left(-\int_z^\infty k_\nu(z') \rho_a(z') dz'\right),$$

where k_ν is the extinction coefficient and ρ_a is the density of the absorbing gas. Consider the special case in which k_ν is independent of height and the absorber has a uniform mass mixing ratio $\mu_a = \rho_a(z)/\rho(z)$, where ρ is the total atmospheric density. Then, using the hydrostatic relation (2.12), the transmittance is found to take the simple form

$$T_\nu(z, \infty) = \exp\left(-\frac{k_\nu \mu_a p}{g}\right) \equiv e^{-\xi_\nu p},$$

say, where

$$\xi_\nu = k_\nu \mu_a / g.$$

Note that $\xi_\nu p$ is the optical depth at frequency ν from the satellite down to pressure p ; see Section 3.6. From equations (7.4) and (7.6) the weighting function is given in terms of p by

$$\mathcal{K} = -p \frac{\partial T_\nu(z, \infty)}{\partial p} = \xi_\nu p e^{-\xi_\nu p}, \quad (7.7)$$

or in terms of Z by

$$\mathcal{K}(Z) = \xi_\nu p_0 \exp(-Z - \xi_\nu p_0 e^{-Z}). \quad (7.8)$$

Note that the weighting function (7.8), like the heating rate in equation (3.32), has a Chapman layer structure: see Figure 7.9. From equation (7.7) it is easy to show that the weighting function takes a maximum value of e^{-1} at a pressure $p = p_m \equiv 1/\xi_\nu = g/(k_\nu \mu_a)$ and at $Z = Z_m = \ln(\xi_\nu p_0) = \ln(k_\nu \mu_a p_0 / g)$, corresponding to unit optical depth. The integral in equation (7.5) therefore tends to be dominated by the Planck function in the neighbourhood of this level and so gives a measure of the average brightness temperature for this region. The thickness of the region depends on how sharply peaked the weighting function is. To investigate this we consider the pressures at which \mathcal{K} equals half its maximum, namely $1/(2e)$: from equation (7.7) these are given by the equation

$$\xi_\nu p e^{-\xi_\nu p} = \frac{1}{2e}.$$

With $\xi_\nu p = x$, this reduces to the equation $2x = \exp(x - 1)$, which is found to have the two roots $x = 0.23$ and 2.68 . The half-maximum points therefore occur at pressures

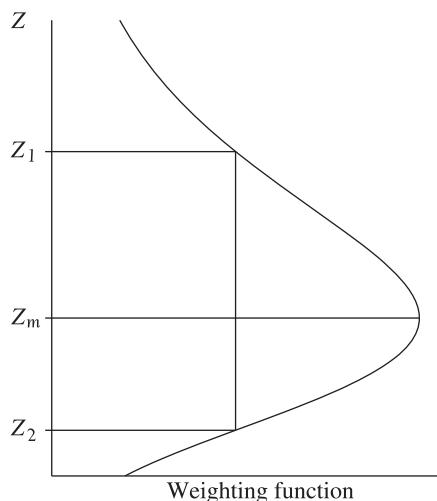


Fig. 7.9

A plot of the weighting function (7.8) as a function of the log-pressure coordinate Z , for an absorber of constant mass mixing ratio and an extinction coefficient that is independent of height, for the case $\xi_v p_0 = 4$, i.e. $p_m = 0.25 p_0 \approx 250 \text{ hPa}$. In this case the maximum of the weighting function is at $Z_m = 1.39$ and the points of half-maximum value are at $Z_1 = 2.85$ (corresponding to $p_1 \approx 58 \text{ hPa}$) and $Z_2 = 0.40$ (corresponding to $p_2 \approx 670 \text{ hPa}$). For other values of p_m the whole plot shifts up or down as a function of Z , without distortion.

$p_1 = 0.23 p_m$ and $p_2 = 2.68 p_m$, corresponding to $Z_1 = Z_m + 1.46$ and $Z_2 = Z_m - 0.99$, and a width in Z of $Z_1 - Z_2 = 2.45$ scale heights, a thickness of about 20 km.

Note that p_m is inversely proportional to the extinction coefficient k_v and that Z_m is linear in $\ln(k_v)$. If measurements of the radiance are made at various frequencies v , corresponding to different k_v , different vertical regions of the atmosphere therefore dominate the integral in equation (7.5) (although the widths of the regions are still 2.45 scale heights). This fact is the basis of a standard method of remote sounding: temperature measurements of different vertical regions of the atmosphere can be inferred from radiance measurements at different frequencies. However, the broad nature of the weighting function is a major disadvantage: temperature measurements averaged over a depth of 20 km are not necessarily very useful. Narrower weighting functions can be obtained by considering cases in which the extinction coefficient is not independent of height (see Problem 7.6) or by using limb rather than nadir sounding, in which radiance measurements are made along a tangent path through the atmosphere, rather than vertically downwards; see Figure 7.1. An example of a stratospheric temperature field, measured by a limb-sounding instrument, is given in Figure 1.8.

In practice, infra-red remote sounding based on measurements from isolated parts of single spectral lines is not feasible, owing to poor signal-to-noise ratios. Therefore averages over many lines must be taken, in such a way that desirable characteristics of the weighting function are preserved.

Infra-red remote sounding of temperature in the Earth's atmosphere invariably uses emission from carbon dioxide, whose mixing ratio is nearly independent of height in the lower and middle atmosphere. Once the temperature profile $T(z)$ has been measured, a

variant of the approach described above allows measurements of composition, with the absorber density ρ_a (which appears in the transmittance T_ν) being regarded as the unknown quantity in equation (7.1).

7.2.2 Backscatter measurements

We now consider a different type of remote-sounding technique for measuring atmospheric composition from space. The idea is to measure solar ultra-violet radiation that has been directed back to space by Rayleigh scattering from atmospheric molecules (see Sections 3.1 and 7.3.2). The simplest case, in which both the incoming solar beam and the scattered beam are vertical, is shown in Figure 7.10. First consider scattering from a layer of thickness dz of atmosphere at height z . If the downward spectral radiance at frequency ν of solar radiation at the top of the atmosphere is $L_{\nu\infty}^\downarrow$, then the downward spectral radiance at height z is

$$L_\nu(z) = L_{\nu\infty}^\downarrow e^{-\chi_\nu(z)}, \quad (7.9)$$

where χ_ν is the optical depth,

$$\chi_\nu(z) = \int_z^\infty k_\nu(z') \rho_a(z') dz' \quad (7.10)$$

(see equation (3.29)), k_ν being the extinction coefficient and ρ_a the absorber density. The radiance scattered upwards from the layer is then of the form

$$L_\nu(z) \rho(z) s_\nu(z) dz$$

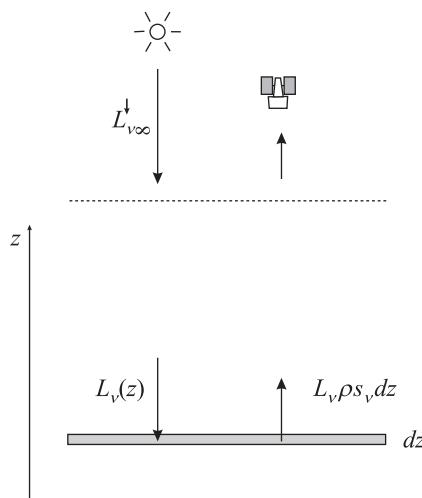


Fig. 7.10

Illustrating the simplest case of backscattering of solar radiation to space, in which the incoming and backscattered beams are vertical. Only the backscattering from the shaded layer, of thickness dz , is indicated. The dashed line represents a notional ‘top of the atmosphere’.

(cf. equation (3.8)), where ρ is the total atmospheric density and s_v is a scattering coefficient. This radiance is attenuated by a further factor $\exp(-\chi_v(z))$ as the scattered beam travels back to space, so that the spectral radiance scattered from this layer that is measured at the satellite is $L_{v\infty}^{\downarrow} \rho(z) s_v(z) dz \exp(-2\chi_v(z))$. Integrating over all layers, we therefore find that the total backscattered spectral radiance measured at the satellite is

$$L_v^s = L_{v\infty}^{\downarrow} \int_0^{\infty} \rho(z) s_v(z) e^{-2\chi_v(z)} dz. \quad (7.11)$$

Now suppose that k_v is independent of z ; then from equation (7.10) the optical depth is

$$\chi_v(z) = k_v \int_z^{\infty} \rho_a(z') dz' = k_v m_a(z), \quad (7.12)$$

where $m_a(z)$ is the total mass of absorber above height z . If the mass mixing ratio μ_a of the absorber were constant, $\bar{\mu}_a$, say, then we would have $m_a(z) = \bar{\mu}_a p/g$, where p is the pressure at height z . Suppose that in fact the mixing ratio is not quite constant, so that

$$m_a(z) = \frac{\bar{\mu}_a p}{g} + m'_a(z), \quad (7.13)$$

where m'_a is small. Then from equations (7.12) and (7.13),

$$e^{-2\chi_v(z)} = e^{-2k_v \bar{\mu}_a p/g} e^{-2k_v m'_a} \approx e^{-2k_v \bar{\mu}_a p/g} (1 - 2k_v m'_a)$$

if $k_v m'_a \ll 1$. If s_v is independent of z and the integration variable is changed from z to p , equation (7.11) becomes

$$L_v^s = \frac{L_{v\infty}^{\downarrow} s_v}{g} \int_0^{p_0} [1 - 2k_v m'_a(p)] e^{-2\tilde{\xi}_v p} dp,$$

where

$$\tilde{\xi}_v = k_v \bar{\mu}_a / g.$$

Then, transforming to the log-pressure variable Z defined in equation (7.4), we get

$$L_v^s = \frac{L_{v\infty}^{\downarrow} s_v}{g} \int_0^{\infty} [1 - 2k_v m'_a(Z)] \mathcal{K}_1(Z) dZ,$$

where

$$\mathcal{K}_1 = p e^{-2\tilde{\xi}_v p} = p_0 \exp[-(Z + 2\tilde{\xi}_v p_0 e^{-Z})]$$

is a weighting function with similar vertical structure to that for emission measurements, given in equations (7.7) and (7.8). The measured radiance L_v^s therefore gives an estimate of the quantity $(1 - 2k_v m'_a)$ near the altitude where the weighting function is a maximum. On making measurements at various frequencies, corresponding to different weighting function peaks, the vertical variation of m'_a and hence of the mixing ratio deviation $\mu_a - \bar{\mu}_a$, can be estimated. An example of an instrument that uses the backscatter technique is the Total Ozone Mapping Spectrometer, which has provided valuable information on the Antarctic ozone hole; see Section 6.7.

7.3 Atmospheric remote sounding from the ground

In this section we consider one example of passive ground-based remote sounding, namely the Dobson ozone spectrophotometer, and two examples of active ground-based remote sounding, namely radars and lidars.

7.3.1 The Dobson ozone spectrophotometer

Ground-based passive remote sounding is usually dominated by the tropospheric signal, so stratospheric properties may be difficult to measure. A notable exception is the amount of ozone, which has its highest concentrations in the stratosphere. This fact is exploited by the Dobson ozone spectrophotometer, which measures the column ozone, the total number of ozone molecules in a column of atmosphere of unit cross-section; see [Section 6.7](#). Consider a case when the Sun is at an angle θ from the vertical; θ is called the **solar zenith angle**. If horizontal variations of the absorber density in the solar beam are negligible, scattering is ignored and the extinction coefficient is independent of position, then the solar spectral radiance at frequency ν at the ground is

$$L_\nu = \tilde{L}_{\nu\infty} e^{-\chi_{\nu 0} \sec \theta} \quad (7.14)$$

(cf. [equation \(7.9\)](#)), where $\tilde{L}_{\nu\infty}$ is the solar spectral radiance along the line of sight from the Sun, at the top of the atmosphere,

$$\chi_{\nu 0} = k_\nu \int_0^\infty \rho_a(z') dz' \equiv k_\nu m_0 \quad (7.15)$$

is the optical path between the ground and space, and m_0 is the total mass of absorber in a vertical column of unit area. The factor $\sec \theta$ in [equation \(7.14\)](#) allows for the increase in path-length when the Sun is not overhead; see [Figure 7.11](#).

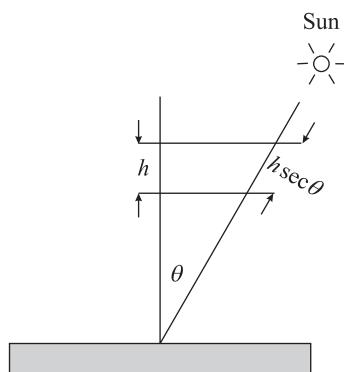


Fig. 7.11

Illustrating how a non-zero solar zenith angle gives a greater path-length for the solar beam. A path h for the overhead Sun becomes $h \sec \theta$ when the Sun is at zenith angle θ . The ground is shown shaded.

From equations (7.14) and (7.15) we get

$$L_\nu = \tilde{L}_{\nu\infty} e^{-k_\nu m_0 \sec \theta},$$

likewise at another frequency ν' but at the same solar zenith angle,

$$L_{\nu'} = \tilde{L}_{\nu'\infty} e^{-k_{\nu'} m_0 \sec \theta}.$$

Therefore, after division and taking logarithms,

$$\ln\left(\frac{L_\nu}{L_{\nu'}}\right) = \ln\left(\frac{\tilde{L}_{\nu\infty}}{\tilde{L}_{\nu'\infty}}\right) + (k_{\nu'} - k_\nu)m_0 \sec \theta. \quad (7.16)$$

The technique used in measuring column ozone is to choose two frequencies in the ultra-violet, one with weak ozone absorption (small k_ν) and one with strong absorption (large $k_{\nu'}$), and carry out measurements at several zenith angles. Then, if the left-hand side of equation (7.16) is plotted against $\sec \theta$, a set of points close to a straight line is found. The slope of the line of best fit is $(k_{\nu'} - k_\nu)m_0$ and, if the extinction coefficients are known, the total mass of absorber (ozone in this case) in a vertical column is obtained. From this the column ozone is readily found.

Dobson exploited this idea in his ozone spectrophotometer, developed during the 1920s and 1930s, using simple but effective experimental techniques, including a null method for measuring $L_\nu/L_{\nu'}$. Essentially the same instrument is still used world-wide for monitoring amounts of column ozone, and it was from a long series of observations with such an instrument that the Antarctic ozone hole was discovered; see Section 6.7.

7.3.2 Radars

Radars employed for atmospheric observation measure the backscatter of pulses of radio waves. Weather radars utilise backscattering of radio waves (with wavelengths of a few centimetres) from water drops and ice crystals in the troposphere to estimate precipitation rates, thus providing important data for weather forecasting systems. In the lower and middle atmosphere, backscattering of waves with a wavelength of a metre or so can take place from inhomogeneities in the atmospheric refractive index \mathcal{N} for radio propagation. These inhomogeneities may, for example, be due to fluctuations of temperature and humidity associated with patches of turbulence, stratification of atmospheric properties in thin layers and regions of ionisation (including ionisation due to meteor trails). The length scales of inhomogeneities that give rise to strong scattering are of the order of half the radio wavelength or greater.³ In this section we shall focus on backscattering from inhomogeneities in \mathcal{N} , rather than from precipitation.

³ Compare with Bragg X-ray diffraction from crystals, for which constructive interference occurs when $2d \sin \theta = n\lambda$, where d is the interplanar spacing, θ is the grazing angle, λ is the wavelength and $n \geq 1$ is an integer. From this equation it follows that $d \geq \lambda/2$. For radar scattering we have a similar formula except that d is interpreted as the length scale of the inhomogeneities in \mathcal{N} .

Suppose that a radar transmits waves in pulses of duration τ , separated by much longer time intervals, and, for simplicity, assume that the pulses are of ‘top-hat’ (‘box-car’) shape; see Figure 7.12(a). In the case of scattering from a single target at a distance, or **range**, r_1 (see Figure 7.12(b)), the scattered pulse detected by the receiver (taken to be at the same location as the transmitter) is of the same duration and shape as the transmitted pulse, if dispersion is neglected. It is clear from Figure 7.12(b) that the time at which the beginning of the pulse is received is $t_1 = 2r_1/c$, where c is the speed of light. However, if scattering takes place at all ranges (see Figure 7.12(c)), then echoes from all ranges between $r_1 - \Delta r$ and $r_1 + \Delta r$ reach the receiver at time t_1 , echoes from all ranges between r_1 and $r_1 + \Delta r$ reach the

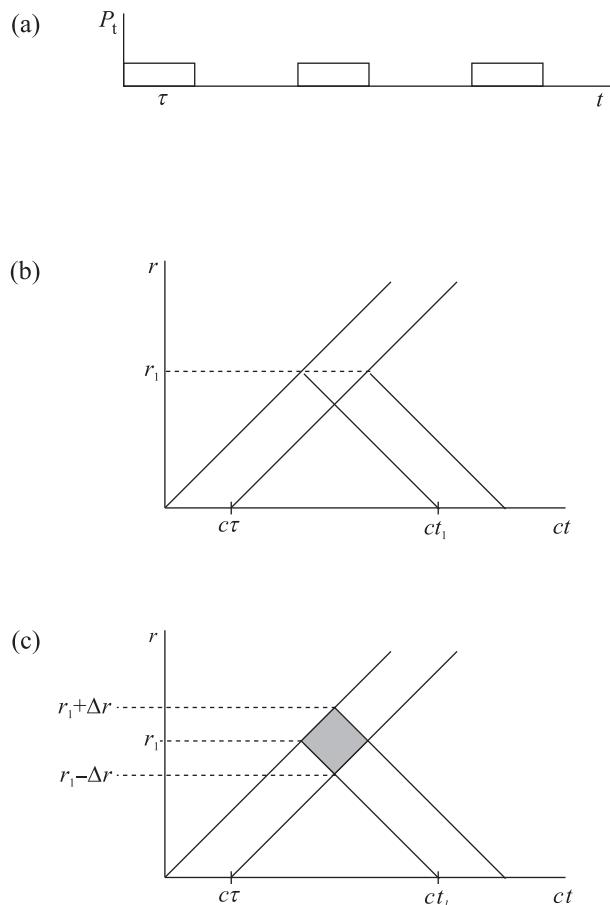


Fig. 7.12

(a) A sketch of the power P_t transmitted by a radar as a function of time, showing the pulses of duration τ , assumed to be of ‘top-hat’ shape. (b) A plot of range r against ct , for the case of a single target at range r_1 , indicating the paths of transmitted and reflected pulses. Because of the scaling of the axes, these paths have slope ± 1 , which makes it clear that $t_1 = 2r_1/c$. (c) The same as (b), but for the case in which scattering takes place at all ranges. The shaded square indicates the time-varying volume of scatterers giving echoes received between times t_1 and $t_1 + \tau$. Since the diagonals of the square are equal, $2\Delta r = c\tau$.

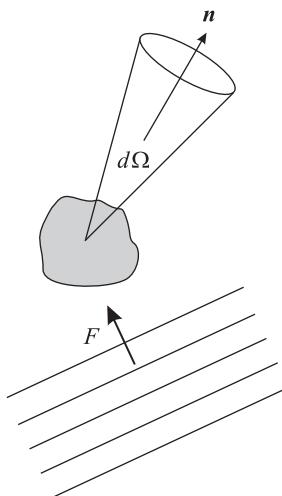


Fig. 7.13 Illustrating the scattering of a plane wave, carrying energy flux F , by an object (shaded). Power $P(\mathbf{n}) d\Omega$ is scattered into solid angle $d\Omega$ in the direction \mathbf{n} .

receiver at time $t_1 + \tau$ and so on, where $2 \Delta r = c\tau$. Therefore, at any instant between t_1 and $t_1 + \tau$ echoes are received from a layer of thickness $\Delta r = c\tau/2$, called the **range resolution**. The range resolution gives the spatial resolution on which atmospheric quantities can be measured: for typical pulse lengths of a few microseconds it is a few hundred metres.

An important concept in scattering theory is the **differential scattering cross-section**, $d\sigma/d\Omega$. Consider scattering from a single object, as in Figure 7.13. A plane wave, carrying an energy flux (power per unit area) F say, is incident on the object and a set of scattered waves, in general propagating outwards in all directions, is produced. Then the differential scattering cross-section is defined as the scattered power $P(\mathbf{n})$ per unit solid angle in a direction \mathbf{n} , divided by the incident flux F :

$$\frac{d\sigma}{d\Omega} = \frac{P(\mathbf{n})}{F}.$$

In the case of electromagnetic waves of frequency ν scattering off an object whose dimensions are much smaller than a wavelength $\lambda = c/\nu$, the differential scattering cross-section satisfies **Rayleigh's law**,

$$\frac{d\sigma}{d\Omega} \propto \frac{1}{\lambda^4} \propto \nu^4; \quad (7.17)$$

proofs of this equation are given in advanced texts.⁴

⁴ Note that equation (7.17) implies that blue light is scattered more than red light. The same result applies to a random distribution of scatterers, which fact was applied by Rayleigh to the scattering of light from atmospheric molecules. Away from the direct solar beam, scattered blue light will predominate, leading to the blue colour of the clear sky. On the other hand, the direct beam itself will appear reddened, since much of the blue spectral component is scattered out; this is especially noticeable at sunset, when the atmospheric path-length of the solar beam is greater than that at midday.

Consider the simple, idealised, case of a radar transmitting power P_t in the form of spherical waves within a cone of solid angle $\alpha < 2\pi$. Suppose that a single isotropic scatterer of constant differential scattering cross-section $d\sigma/d\Omega$ is present at range r on the axis of the cone; the energy flux F incident on the scatterer is $P_t/(\alpha r^2)$. If the radar receiver has an antenna area A , perpendicular to the axis of the cone, it subtends a solid angle A/r^2 at the scatterer. It follows that the scattered power received by the radar is

$$P_r = \frac{P_t}{\alpha} \left(\frac{d\sigma}{d\Omega} \right) \frac{A}{r^4}. \quad (7.18)$$

(The transmittance at radio frequencies may be taken to be unity.) Note the r^{-4} dependence of P_r ; this radial dependence would apply for scattering from an aircraft, for example.

Now suppose instead that the beam is filled with n scatterers per unit volume, each of differential scattering cross-section $d\sigma/d\Omega$, between ranges r and $r + \Delta r$, where $\Delta r = c\tau/2$ is the range resolution, τ being the pulse length. The beam need not be vertical; if it is pointed at a zenith angle θ it can be shown from equation (7.18) that the received power P_r is given in terms of the transmitted power P_t by

$$P_r = P_t \frac{A \cos \theta}{r^2} \beta \eta \Delta r, \quad (7.19)$$

where A is the area of the radar array, β is a volume scattering coefficient and η is an efficiency and geometric factor for the radar. If the scattering is due to turbulence, the scattering coefficient β may depend in quite a complicated way on turbulent fluctuations of the refractive index \mathcal{N} . If the scattering is from random thermal motion of free electrons in the ionised part of the atmosphere above about 60 km altitude, then $\beta \approx n_e \sigma_e / 2$, where n_e is the free-electron number density and σ_e is the backscattering cross-section of an electron. Note that the radial dependence in equation (7.19) is r^{-2} , in contrast to the r^{-4} dependence in equation (7.18); this is because the scattering volume is itself proportional to r^2 .

An alternative situation is one in which vertically emitted radio waves are reflected from an extensive horizontal layer in which the refractive index varies rapidly with height. In this case

$$P_r = P_t \frac{A^2}{r^2} \frac{\eta'}{\lambda^2} \mathcal{R},$$

where η' is another efficiency and geometric factor, \mathcal{R} is the power reflection coefficient for the layer and λ is the radio wavelength. This is an example of **partial** or **specular** reflection. The dependence of \mathcal{R} on the vertical gradient of the refractive index can be calculated in some cases, the simplest being that in which there is a jump $\Delta \mathcal{N}$ at a given height; see Problem 7.9.

A semi-empirical expression for the refractive index for radio waves of angular frequency ω is

$$\mathcal{N} - 1 = a_1 \frac{e}{T^2} + a_2 \frac{p}{T} - \frac{n_e q_e^2}{2 \epsilon_0 m_e \omega^2}, \quad (7.20)$$

where p is the atmospheric pressure, e is the partial pressure of water vapour, T is the temperature, n_e is the number density of free electrons, q_e is the electronic charge, m_e is the mass of the electron, ϵ_0 is the permittivity of free space, and a_1 and a_2 are constants.

The first term on the right-hand side of equation (7.20) results from bound electrons present in density fluctuations of water vapour: this is most important in the lower troposphere, where water-vapour mixing ratios are highest. The second term (which is proportional to the air density, by equation (2.2)) results from bound electrons present in density fluctuations of dry air and is most important in the upper troposphere and the stratosphere. The third term results from free electrons and is most important above about 50 km, where the free electron density increases rapidly with height. Each of the terms on the right-hand side of equation (7.20) is much less than 1 in magnitude.

Partial reflection techniques allow the radio refractive index \mathcal{N} to be measured and hence offer the possibility of estimating the atmospheric quantities on which \mathcal{N} depends, such as temperature and the free-electron density.

By measuring the Doppler shift of radar echoes from turbulence, an estimate of the radial (line-of-sight) component V of the wind velocity can be obtained, assuming that the turbulent region moves with the mean wind. Consider two short radar pulses emitted at times $t = 0$ and $t = t_0$, respectively. If the first pulse reaches the scattering region at a radial distance r_1 , it is received back at the radar at time $2r_1/c$; see Figure 7.14. By the time the second pulse has reached the scattering region, this region has moved. If it is now at a range r_2 , consideration of the extra distance travelled by the outgoing radar wave shows that $r_2 = r_1 + Vt_0 + V(r_2 - r_1)/c$. However, since $V \ll c$ (this calculation is of course non-relativistic) we have $r_2 = r_1 + Vt_0$ to a very good approximation. The time difference between the reception of the reflected pulses is therefore $t_0 + 2(r_2 - r_1)/c = t_0(1 + 2V/c)$.

If we apply the same argument to two adjacent peaks, say, of a sinusoidal radio wave of angular frequency $\omega_0 = 2\pi/t_0$, then the Doppler-shifted angular frequency of the reflected wave is $\omega_r = \omega_0/(1 + 2V/c) \approx \omega_0(1 - 2V/c)$. Hence the change of frequency is

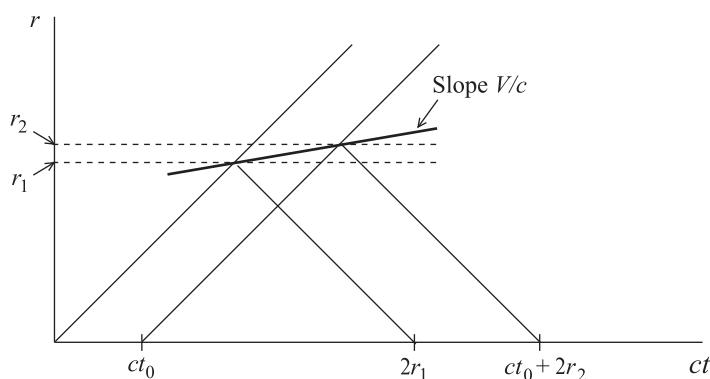


Fig. 7.14

A plot of radial distance r against ct , illustrating the Doppler effect for radar measurements. The paths of two transmitted pulses and their reflections are shown. The bold sloping line shows the path of the scattering region; its slope is greatly exaggerated, since $V \ll c$. The Doppler effect is demonstrated by the fact that the time difference between the returning pulses is greater than that between the transmitted pulses. See the text for further details.

$\delta\omega = \omega_r - \omega_0 = -2\omega_0 V/c$ and, if $\delta\omega$ is measured, the radial component V of velocity is obtained.

If the radar is pointed in a direction given by the unit vector \mathbf{n} , then $V = \mathbf{n} \cdot \mathbf{u}$, where \mathbf{u} is the wind vector. By varying the direction \mathbf{n} and assuming that \mathbf{u} does not vary significantly between the different scattering regions thus observed, the separate horizontal and vertical components of the wind can be inferred.

A variant on the above method is to measure the Doppler shift of radio waves scattered by patches of ionisation produced by meteors entering the atmosphere. These give wind estimates at altitudes between 80 and 100 km.

Another method for measuring the horizontal components of velocity of a moving scattering region is the spaced antenna technique. As the scattering region moves through the transmitted radar beam, it produces a moving diffraction pattern at the ground, which is sampled by a number of spaced receivers. Cross-correlation of the received signals allows the time delays between each of the receivers, and hence the required velocity, to be measured. As with the Doppler method, it is necessary to assume that the scattering region is moving with the wind.

Wind measurements of these types have been very important in determining the structure of gravity waves (see [Section 5.4](#)) in many parts of the atmosphere. An example of the use of the Doppler technique is the observation by [Balsley et al. \(1983\)](#) of an inertia–gravity wave, shown in [Figure 1.7](#). The magnitude of turbulent fluctuations about the mean wind can be obtained from the Doppler width of the echoes.

7.3.3 Lidars

A lidar⁵ consists of a vertically directed laser that transmits pulses of radiation, typically at visible and ultra-violet wavelengths, into the atmosphere; the backscattered radiation is measured, after a time delay, at the same site as the transmitter. The geometric and physical principles involved are similar to those for measurement using radars, described in [Section 7.3.2](#), except that the effects of atmospheric absorption must now be considered and scattering takes place from molecules and aerosols, rather than from refractive-index variations. If a pulse of length τ and power P_t is transmitted by the laser, then the power scattered from a layer of thickness Δz at height z and detected by the receiver is

$$P_r = P_t e^{-2\chi(z)} n(z) \left(\frac{d\sigma}{d\Omega} \right) \frac{A}{z^2} \eta \Delta z.$$

Here $\chi(z)$ is the optical path (at the frequency ν of the laser) between the ground and height z ; $n(z)$ is the number density of the scattering species and $d\sigma/d\Omega$ is its differential

⁵ Light detection and ranging.

backscattering cross-section; and A is the telescope area and η the optical efficiency of the receiver. The height z at which scattering takes place can be inferred from the time delay $2z/c$ between transmission and reception of the pulse, where c is the speed of light. The range resolution $\Delta z = c\tau/2$ typically lies between a few tens of metres and a hundred metres.

In the case in which the scattering is Rayleigh scattering from any atmospheric molecule, $d\sigma/d\Omega = av^4$ by equation (7.17), where a is a known constant. The method can then be used to obtain the total number density $n(z)$ of the atmosphere and hence the density $\rho(z) = \bar{m}n(z)$ (where \bar{m} is the mean molecular mass), at height z . Rayleigh scattering is important for wavelengths less than about 3 μm and allows measurements of density from heights between about 10 and 90 km. The temperature can be obtained by use of the ideal gas equation (2.2) and the hydrostatic equation (2.12); see Problem 7.10.

In the case of scattering from aerosol particles, whose sizes may approach or exceed the laser wavelength, Rayleigh scattering does not apply, and the more complex Mie scattering is the relevant process. It is difficult to use lidar measurements to quantify aerosol concentrations, but one can instead plot the lidar backscatter ratio R , defined as the ratio of the total number of backscattered photons (from molecules and aerosols) at frequency ν to the number of backscattered photons from molecules alone:

$$R = \frac{\beta_{\text{mol}} + \beta_{\text{aerosol}}}{\beta_{\text{mol}}},$$

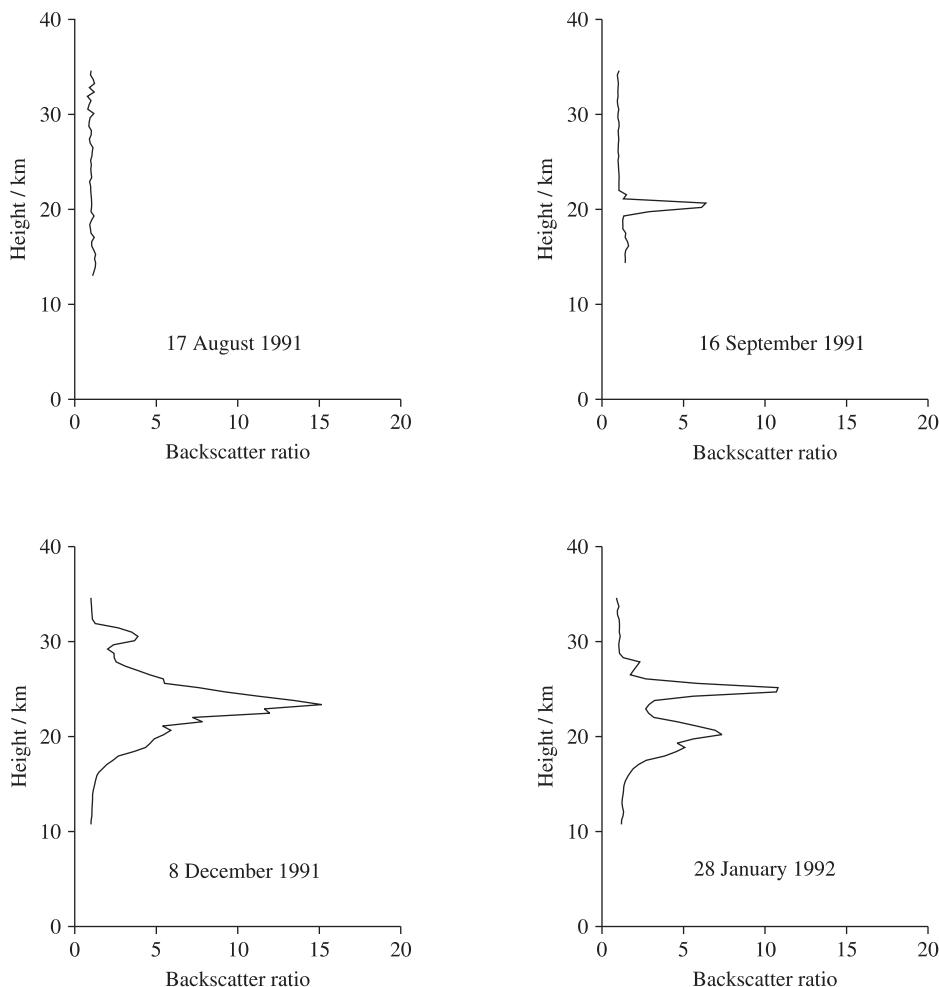
where $\beta_i = n_i d\sigma_i/d\Omega$. Layers of high aerosol content, due for example to volcanic eruption, show up as sharp peaks in $R(z)$; see Figure 7.15.

Lidars with tunable dye lasers, tuned to the frequency of specific atomic or molecular transitions, can be used to measure the vertical distribution of chemical species, including various alkali atoms (deposited in the atmosphere from meteors) and ozone. For example the ${}^2\text{S}_{1/2} - {}^2\text{P}_{3/2}$ transition of sodium at 589.0 nm is associated with an enhanced scattering cross-section (this is an example of **resonance scattering**), which allows the lidar measurement of amounts of sodium at heights of 80–100 km.

Further reading

A brief discussion of remote sounding from space is given by Houghton (2002), and comprehensive treatments by Houghton *et al.* (1984) and Hanel *et al.* (2003). A description of the Dobson ozone spectrometer is given by Dobson and Normand (1957). The basic physics of atmospheric radars and some typical results are presented by Balsley and Gage (1980); for a review of lidars see Thomas (1987).

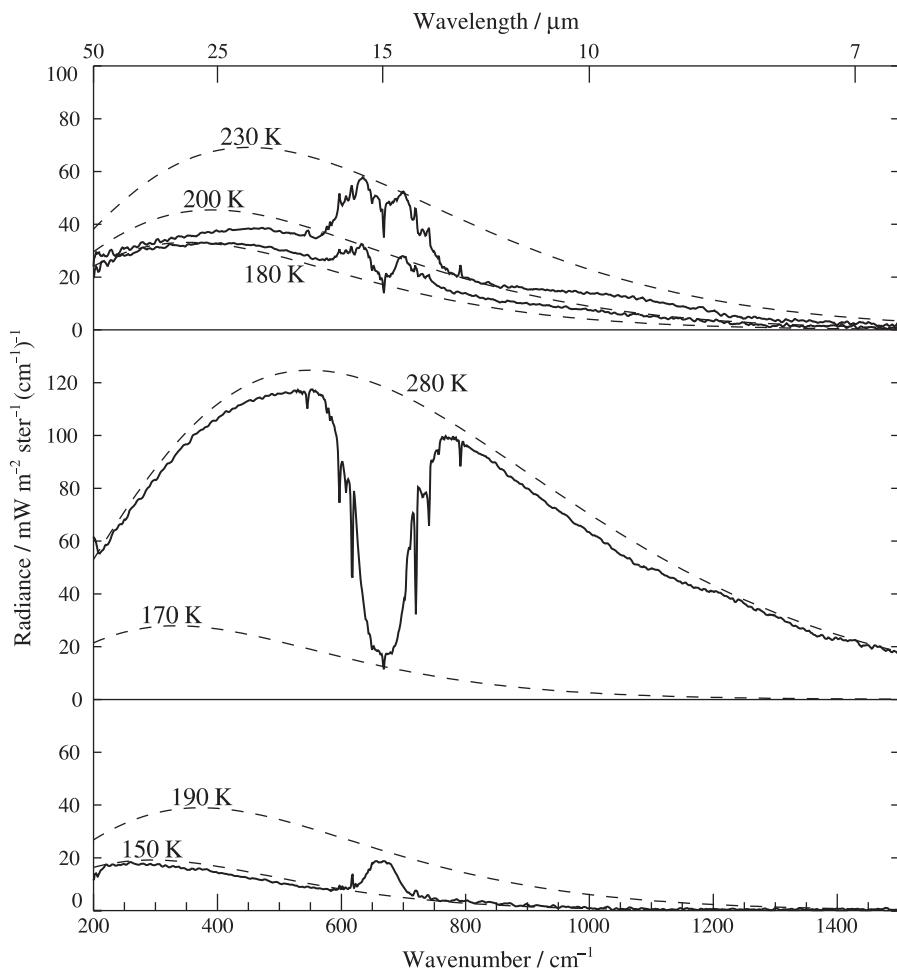
Basic texts on electromagnetic waves include those by Lorrain *et al.* (1988) and Grant and Phillips (1990). The detailed physics of Rayleigh scattering is covered in the advanced texts by Jackson (1998) and Born and Wolf (1980); a descriptive account of the phenomenon is given by Lynch and Livingston (2001), who also detail many other examples of atmospheric optical phenomena. Mie scattering is treated by Liou (2002).

**Fig. 7.15**

Height profiles of the lidar backscatter ratio R measured at Aberystwyth (52°N , 4°W), showing a build-up in the lower stratosphere of volcanic aerosols from Mt Pinatubo (see Vaughan *et al.* (1994)). On 17 August 1991, $R \approx 1$ throughout the profile, indicating that little aerosol is present. On 16 September, R exhibits a peak in a thin layer at about 20 km altitude. On 8 December there is a very large peak, extending over a deep layer centred at about 23 km. On 28 January 1992 a double-peaked structure is present. Based on data provided by Dr G. Vaughan.

Problems

Problem 7.1 Consider a thin layer of gas of uniform density, between altitudes z and $z + d$ in a spherical atmosphere, emitting thermal radiation. Show that the ratio of the maximum mass of the gas in a limb-viewing path to the mass of gas in a nadir-viewing path is approximately $(8a/d)^{1/2}$, where a is the Earth's radius. Evaluate this ratio for $d = 10 \text{ km}$ and comment on your result.

**Fig. 7.16**

Emission spectra from Mars obtained with the IRIS instrument on Mariner 9 (adapted from Hanel *et al.* (1972)). Top panel: spectra recorded over the south polar region; upper curve includes a smaller fraction of the polar ice cap than the lower curve. Middle panel: spectrum recorded near 21° S. Lower panel: spectrum recorded near 66° N; note that the condensation temperature of CO₂ at Martian surface pressures is about 145 K. Diagram prepared with the help of Dr S. R. Lewis, using data from the Planetary Data System.

Problem 7.2 Derive equation (7.1) from equation (3.13), assuming local thermodynamic equilibrium, given that the Earth's surface is a black body.

Problem 7.3 Figure 7.16 shows emission spectra from Mars.

1. What do the fine features between 50 and 25 μm, especially in the south polar spectra, indicate?
2. What is responsible for the large feature centred on 15 μm?

3. What can be said about the surface and atmospheric temperatures?
4. What can be learnt from a comparison of terrestrial (e.g. Figure 7.8) and Martian midlatitude spectra?

Problem 7.4 Measurements of outgoing radiation are made at three wavelengths by a nadir-viewing radiometer in Earth orbit, with the following results:

1. brightness temperature 310 K (at 11 μm , in the atmospheric window),
2. brightness temperature 220 K (at 15 μm , in the CO_2 band), and
3. brightness temperature 280 K (at 9.6 μm , in the O_3 band).

Assuming that the surface is black, that the CO_2 absorption is so strong at wavelength (b) that the stratosphere is black there and that all the ozone resides in the stratosphere, calculate the transmittance of the ozone layer at 9.6 μm . What could we learn about the ozone layer from this?

(The Planck function B_ν at 9.6 μm is approximately 5, 21 and 36 in units of $10^{-13} \text{ W m}^{-2} \text{ steradian}^{-1} \text{ Hz}^{-1}$ at 220, 280 and 310 K, respectively.)

Problem 7.5 The 11 μm region is often referred to as being in a ‘window’ in the terrestrial atmosphere, but there is still some absorption, mainly due to water vapour. Assuming that the absorption coefficient k due to the water is $k = \alpha e$, where e is the vapour pressure and α is a constant, and that the mass mixing ratio μ at pressure level p is $\mu_0(p/p_0)^3$, where μ_0 and p_0 refer to the bottom boundary and μ_0 is small, show that the transmittance of a vertical path from level p to a satellite is $\exp[-\beta(p/p_0)^8]$ where

$$\beta = \frac{\alpha}{g} \frac{M}{M_w} \frac{\mu_0^2 p_0^2}{8}$$

and M_w and M are the molar masses of water vapour and air, respectively.

Given that $\alpha = 7 \times 10^{-6} \text{ m}^2 \text{ kg}^{-1} \text{ Pa}^{-1}$ and that the temperature T_0 and relative humidity A_0 at the bottom boundary are 290 K and 70%, respectively, show that $\beta = 0.102$ and calculate the transmittance from the surface to the satellite. (The SVP of water vapour at 290 K is 1936.7 Pa.)

Write down an expression for the net deficit in the spectral radiance received from the surface by a nadir-viewing radiometer due to attenuation and emission by water vapour. Evaluate this deficit for the case of a (black) sea surface at temperature T_0 for the conditions given above, assuming that $T = T_0(p/p_0)^\delta$ and $B_\nu = \eta(T/300)^\gamma$. Take $p_0 = 10^5 \text{ Pa}$, $\delta = 0.29$, $\eta = 4 \times 10^{-12} \text{ W m}^{-2} \text{ steradian}^{-1} \text{ Hz}^{-1}$ and $\gamma = 4.5$.

$$\left(\int_0^1 y^{8.305} \exp(-0.102y^8) dy = 0.102 \right)$$

What would be the error in inferring the temperature of the sea surface from radiance measurements if the effect of the water vapour were not taken into account? Is this error significant?

Problem 7.6 What is meant by a **weighting function** in atmospheric remote sounding? Show that, if the transmittance $\mathcal{T}(p) = \exp(-\beta p^\alpha)$, where α and β are constants, and if the height-like variable $Z = \ln(p_0/p)$ is used as a vertical coordinate, the vertical weighting function is

$$\mathcal{K}(p) = \alpha\beta p^\alpha \exp(-\beta p^\alpha).$$

Find α and β for an isothermal atmosphere containing an absorber with a small and constant mass mixing ratio μ when

1. the absorber is grey with constant absorption coefficient k ;
2. the radiometer is sensitive to a single frequency in the far wings of a Lorentz line; and
3. the radiometer is sensitive to a spectral interval in which the transmittance from level p to the satellite is given by the strong limit of the Goody random model,

$$\mathcal{T} = \exp\left(-\frac{2}{\delta}(Sm\gamma_L)^{1/2}\right),$$

where S is the mean line strength, m is the mass of absorber per unit area in the path, γ_L is the mean Lorentz half-width and δ is the mean line spacing.

Problem 7.7 Find the pressure levels at which the peaks of the weighting functions occur and find also the widths (in terms of pressure) at half maximum of the weighting functions of the previous question. Sketch the functions for cases (a) and (b).

(The roots of $2x = e^{x-1}$ are approximately 2.68 and 0.23.)

Problem 7.8 Verify the last term on the right-hand side of equation (7.20) for the radio refractive index, by ignoring the effects of bound electrons and proceeding as follows.

It can be shown (see Lorrain *et al.* (1988) or Jackson (1998)) that the dispersion relation for transverse plasma waves is

$$\omega^2 = \omega_p^2 + c^2 k^2,$$

where the **plasma frequency** ω_p is given by

$$\omega_p^2 = \frac{n_e q_e^2}{\epsilon_0 m_e},$$

in the notation of Section 7.3.2. Hence show that, if $\omega^2 \gg \omega_p^2$,

$$\mathcal{N} \equiv \frac{ck}{\omega} \approx 1 - \frac{n_e q_e^2}{2\epsilon_0 m_e \omega^2}.$$

Problem 7.9 Suppose that the atmosphere has a radio refractive index $\mathcal{N}(z)$ given by

$$\mathcal{N} = \mathcal{N}_1 \quad \text{for } z < z_0, \quad \mathcal{N} = \mathcal{N}_2 \quad \text{for } z > z_0.$$

Consider a vertically propagating electromagnetic wave incident on the level $z = z_0$ from below. By using standard methods for calculating the reflected wave (see, e.g., Grant and

Phillips (1990) or Lorrain *et al.* (1988)) show that the power reflection coefficient \mathcal{R} (the ratio of the reflected to the incident power) is given by

$$\mathcal{R} = \frac{1}{4} (\mathcal{N}_1 - \mathcal{N}_2)^2$$

when \mathcal{N}_1 and \mathcal{N}_2 are both close to 1. (You may find it convenient to take $z_0 = 0$ for this calculation.)

Problem 7.10 Given observations of density $\rho(z)$ from lidar measurements and the temperature $T(z_m)$ at a single height z_m , use the ideal gas equation and hydrostatic balance to show that the full temperature profile $T(z)$ can be obtained from

$$T(z) = \frac{1}{\rho(z)} \left(\rho(z_m) T(z_m) + \frac{g}{R_a} \int_z^{z_m} \rho(z') dz' \right).$$

Show that the term $\rho(z_m) T(z_m) \rightarrow 0$ as $z_m \rightarrow \infty$. (In practice z_m is taken as the maximum height of the lidar measurements and a value of $T(z_m)$ is guessed; the resulting values of $T(z)$ are not very sensitive to this guess.)

This chapter presents a selection of topics on the physics of climate change. By way of introduction, in [Section 8.1](#) we briefly discuss greenhouse gases and the radiative forcing associated with several drivers of climate change. Then in [Sections 8.2](#) and [8.3](#) we introduce a very simple time-dependent, ‘energy balance’, climate model. This model illustrates some key concepts that arise in the study of the physics of the Earth’s climate and its response to external forcing, and in the diagnosis of the highly complex models that are used to simulate the climate of the past and present and to predict future climate. In [Section 8.4](#) we examine some elementary aspects of the important topic of climate feedbacks. Finally, in [Section 8.5](#) we use another simple model to examine the basic physics of the process by which the radiative forcing due to carbon dioxide increases with its concentration.

Our emphasis in this chapter is on the underlying physical principles of climate change; we shall not discuss in any detail the climate-change projections by the current range of complex general circulation models. Comprehensive information on these projections is provided for example by the Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC). The most recent such report is that of [Solomon *et al.* \(2007\)](#), which includes useful summaries for non-specialists and a glossary of technical terms.

8.1 Introduction

The **greenhouse effect** was introduced in [Section 1.3](#) and discussed further in [Section 3.7](#). We now consider the effects of different **greenhouse gases**, i.e. gases that absorb and emit infra-red radiation but allow solar radiation to pass through without significant absorption.

As a starting point, consider a hypothetical atmosphere containing no greenhouse gases or other absorbers. Given an unreflected (i.e. absorbed) solar irradiance $F_0 \approx 240 \text{ W m}^{-2}$, the surface temperature (assumed uniform) is about 255 K, as shown in [Section 1.3.1](#). To quantify the effects of greenhouse gases in raising this to the current global-mean surface temperature of about 288 K, which has a black-body irradiance of about 390 W m^{-2} , one can say that these gases currently ‘provide a greenhouse effect’ of about $390 - 240 = 150 \text{ W m}^{-2}$. (We ignore the long-wave effects of clouds here.) In physical terms, the greenhouse gases absorb some of the radiation emitted by the warm surface and emit a lesser amount of radiation from cooler tropospheric regions aloft; they also emit energy downwards, and hence help to warm the surface. See [Problem 8.1](#) for a simple model of this process.

An important corollary is that, all else being equal, a decrease of upper-tropospheric temperatures leads to an increase in the greenhouse effect.

It would be of interest to apportion this greenhouse effect between the different gases; however, this cannot be done in a simple way, given the large difference in composition between the hypothetical and actual atmospheres, because the effects of the separate gases do not add linearly. (See [Problem 8.2](#).) For example, numerical experiments in which different greenhouse gases are in turn completely removed from the atmosphere show that the greenhouse effect due to both CO₂ and water vapour is smaller than the sum of the individual effects of each of them. Nevertheless it is clear that water vapour is the most important greenhouse gas in determining the current atmospheric state, contributing around 2/3 of the current greenhouse effect, and that CO₂ is the next most important, contributing around 1/4.

Different considerations apply when we consider the response of any given climate state (for example, the current climate or the ‘pre-industrial’ climate), to *perturbations* to the various absorbing gases or other climate variables. Examples of interest include perturbations to greenhouse gas concentrations, aerosols, clouds or albedo, and especially those perturbations caused by human activities. Here an important quantity is the **radiative forcing**, defined as the net decrease of the upward irradiance at the tropopause due to a change in some climate-change driver, such as a specified increase in the amount of a gaseous absorber. It is assumed that the surface and troposphere are otherwise held fixed and that stratospheric temperatures, if perturbed, have readjusted to radiative-dynamical equilibrium.

The net upward irradiance at the tropopause equals the outgoing irradiance minus the unreflected incoming solar irradiance there; in equilibrium, this would be zero. If the radiative forcing (RF) is positive, the incoming irradiance exceeds the outgoing irradiance and there is a positive net heat flux into the climate system below the tropopause, leading to a warming of the climate; a negative RF contributes to a cooling of the climate. An important example of a positive RF is provided by the **enhanced greenhouse effect**, in which an increase in a tropospheric greenhouse gas decreases the outgoing long-wave irradiance (often called the **outgoing long-wave radiation**, or OLR). Calculation of the RF due to atmospheric aerosols is very complex: they lead to ‘direct’ effects, by scattering and absorbing long-wave and short-wave radiation, and to ‘indirect’ effects, by modifying the microphysical and hence the radiative properties of clouds; overall, both effects are believed to give negative RFs.

[Figure 8.1](#) shows estimated values of the RFs associated with the main climate-change drivers for the period 1750–2005; for the comparatively small perturbations involved, the separate effects do add fairly linearly. Drivers with positive RFs include the long-lived greenhouse gases carbon dioxide, methane (CH₄), nitrous oxide (N₂O) and the halocarbons (including CFCs), and also tropospheric ozone and stratospheric water vapour. Aerosols have negative RFs, while surface albedo changes can contribute either positive or negative RFs. These drivers are all significantly affected by human activities; the only significant and sustained natural RF over the period was due to an increase in solar irradiance, although the 11-year solar cycle causes oscillatory changes in RF, and stratospheric aerosols resulting from massive volcanic eruptions can give short-term negative RFs.

Radiative forcing of climate between 1750 and 2005

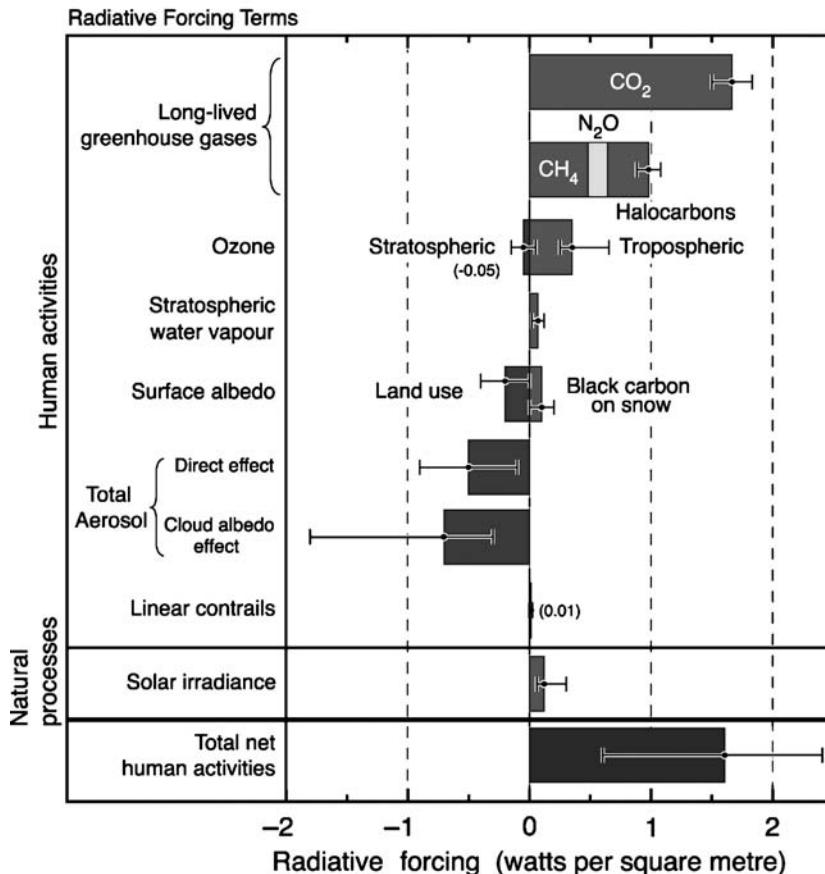


Fig. 8.1

Summary of the principal components of the radiative forcing of climate change. The values represent the forcings in 2005 relative to the start of the industrial era, taken as 1750. The thin black line attached to each bar represents the range of uncertainty for the respective value. For the CH₄, N₂O and halocarbon block, the error is for the three gases combined. The levels of scientific understanding of the various components are estimated as high for the long-lived greenhouse gases, medium for ozone and medium-low or low for the others. (Figure from Solomon *et al.* (2007), FAQ 2.1, Figure 2.)

Tropospheric water vapour does not feature in Figure 8.1: rather than providing an RF of its own, water vapour primarily acts by modifying the effects of other climate drivers. This is because it is a short-lived gas: its concentration varies on small time and space scales, mainly by exchange with the large liquid water reservoir through rapid condensation and evaporation processes, which are strongly temperature-dependent. Since a temperature increase leads to an increase in the amount of water vapour by the Clausius–Clapeyron equation (2.37), and increased water vapour gives a decreased OLR and hence a warmer climate, water vapour provides a **positive feedback** on temperature changes: see Section 8.4.2.

8.2 An energy balance model

In this section we introduce a very simple zero-dimensional energy balance model (EBM) of the time-dependent response of the climate system to radiative forcing. In this model the term ‘climate system’ is taken to encompass the top 100 m or so of the ocean (the ocean’s ‘mixed layer’) together with the troposphere and the land surface: these three components are assumed to be represented by a ‘global-mean’ temperature T . Heat transport into the deep ocean is neglected for simplicity, so the only heat transfers into and out of the climate system are due to the absorbed solar (short-wave) radiation and the outgoing long-wave radiation, respectively.

This is clearly an extremely simplified version of the real, highly complex, climate system. However, it allows us to identify some important basic properties of the real system, and to explore some possible ways in which the climate may change in response to radiative forcing.

The EBM is described by the equation

$$C \frac{dT}{dt} = F^\downarrow - F^\uparrow, \quad (8.1)$$

which states that, averaged over the Earth, the rate of change of heat content per unit horizontal area of the climate system is given by the net heat flux into the climate system, i.e. the difference between the incoming and outgoing radiative power per unit area. Here F^\downarrow is the solar (short-wave) irradiance absorbed by the climate system, and is equal to the unreflected incoming solar irradiance at the top of the troposphere. Neglecting any absorption of solar radiation above the tropopause we can take

$$F^\downarrow = F_0 \equiv \frac{1}{4}(1 - A)F_s, \quad (8.2)$$

(see [equation \(1.3\)](#)), where $F_0 \approx 240 \text{ W m}^{-2}$, A is the planetary albedo and F_s is the total solar irradiance (TSI). Note that this expression for F^\downarrow does not depend on the temperature or greenhouse gas concentration of the atmosphere. On the other hand, the outgoing (long-wave) irradiance or OLR F^\uparrow does generally depend on the temperature, water vapour content, carbon dioxide concentration, etc., of the troposphere.

The constant C is the heat capacity per unit horizontal area of the climate system. A simple approximation is to take it as the heat capacity per unit area of the oceanic mixed layer times the fraction (~ 0.7) of the Earth’s surface covered by ocean.

We suppose for the moment that the net heat flux $F^\downarrow - F^\uparrow$ into the climate system depends only on the temperature T and the concentration U , say, of a particular greenhouse gas,

$$F^\downarrow - F^\uparrow = Q(T, U). \quad (8.3)$$

In a steady-state climate at constant temperature $T = T_{ss}$ and concentration $U = U_{ss}$, we have $CdT/dt = 0$ and so

$$Q(T_{ss}, U_{ss}) = 0. \quad (8.4)$$

If this steady state is now slightly perturbed, so that $T = T_{ss} + T'(t)$ and $U = U_{ss} + U'(t)$, where $T' \ll T_{ss}$ and $U' \ll U_{ss}$, we obtain the linearised form

$$\begin{aligned} C \frac{dT'}{dt} &= Q(T_{ss} + T', U_{ss} + U') \\ &\approx \frac{\partial Q}{\partial T} T' + \frac{\partial Q}{\partial U} U', \end{aligned} \quad (8.5)$$

where we have expanded $Q(T, U)$ to first order in T' and U' and used equation (8.4). The partial derivatives are evaluated at $(T, U) = (T_{ss}, U_{ss})$.

We now define the **climate feedback parameter**

$$\alpha = -\frac{\partial Q}{\partial T}. \quad (8.6)$$

An analogy with feedback in a simple electronic circuit can be made by considering an electromotive force \mathcal{E} across an inductance L that carries current \mathcal{I} , so that $\mathcal{E} = Ld\mathcal{I}/dt$. A feedback is present if \mathcal{E} depends on \mathcal{I} , for example through the linear relation $\mathcal{E} = \mathcal{F} - \alpha\mathcal{I}$ where \mathcal{F} and α are constants.

The climate feedback parameter α quantifies how the net downward heat flux Q varies with temperature. If Q decreases with temperature, α is positive and tends to damp temperature perturbations. On the other hand if Q increases with temperature, α is negative and tends to enhance temperature perturbations.

In Section 8.4 we examine in more detail the processes that contribute to climate feedback. As noted there, typical global-mean values of α are positive under current climate conditions, though under extreme climates they might become negative.

We also define the **radiative forcing**

$$F = \frac{\partial Q}{\partial U} U' \quad (8.7)$$

so that equation (8.5) can be written

$$C \frac{dT'}{dt} + \alpha T' = F(t). \quad (8.8)$$

The radiative forcing F as defined in equation (8.7) quantifies the change in the net inward heat flux $F^\downarrow - F^\uparrow$ at the top of the troposphere associated with a small change U' of a greenhouse gas from its steady-state value. As we have seen in Section 8.1, it can be defined more generally for a finite change and applied to quantities other than greenhouse gases.

It is not possible to give precise values of the feedback parameter or the heat capacity for the Earth's climate system. However, experiments with the complex climate models cited in Solomon *et al.* (2007) suggest that $\alpha \approx 1 \text{ W m}^{-2} \text{ K}^{-1}$ and $C \approx 1 \text{ GJ K}^{-1} \text{ m}^{-2}$ might be typical order-of-magnitude estimates.

It should be noted that our linearised EBM (8.8) ignores the possibility of climate change in the absence of an external forcing; this might be brought about, for example, by non-linear interactions between different components of the climate system.

8.3 Some solutions of the linearised energy balance model

Equation (8.8) is a first-order linear differential equation for the deviation T' of the global-mean temperature from its steady-state value T_{ss} . A formal solution of this equation, for arbitrary radiative forcing $F(t)$ and suitable initial conditions, can easily be obtained by the standard ‘integrating factor’ method, as outlined below. In this section we assume that $\alpha > 0$.

Note that the quantity

$$\tau = \frac{C}{\alpha} \quad (8.9)$$

has the dimensions of time: it is in fact a relaxation time, but is also called the **feedback response time** (FRT). Using the order-of-magnitude estimates of C and α quoted in the previous section, we expect τ to be roughly on the order of 10^9 s \approx 30 years for the Earth’s climate system.

To solve equation (8.8) we divide by C and multiply by the integrating factor $\exp(t/\tau)$ to get

$$\frac{d}{dt} (T' e^{t/\tau}) = \frac{F(t)}{C} e^{t/\tau}. \quad (8.10)$$

We assume that $T' = 0$ at an initial time $t = 0$, and integrate equation (8.10) to obtain the formal solution

$$T'(t) = \frac{e^{-t/\tau}}{C} \int_0^t F(u) e^{u/\tau} du. \quad (8.11)$$

To get a feel for the behaviour of solutions of the form (8.11), we look at the responses to a few simple but important examples of forcings $F(t)$ with different time variations.

Step function forcing

Here we take F to be a step function of time:

$$F(t) = 0 \quad \text{for } t \leq 0, \quad = F_1 \quad \text{for } t > 0, \quad (8.12)$$

where F_1 is a constant. It is then easy to show that the temperature response is given by

$$T'(t) = S (1 - e^{-t/\tau}) \quad \text{for } t > 0, \quad (8.13)$$

where

$$S = \frac{F_1}{\alpha} \quad (8.14)$$

is called the **climate sensitivity**:¹ it is the asymptotic value approached by the temperature perturbation T' as $t \rightarrow \infty$, when $\alpha T' \sim F$ in equation (8.8). Figure 8.2(a) plots the time evolution of the step-function forcing and the resulting temperature response. Note that T'

¹ Confusingly, α^{-1} is also sometimes called the climate sensitivity; we shall avoid this usage.

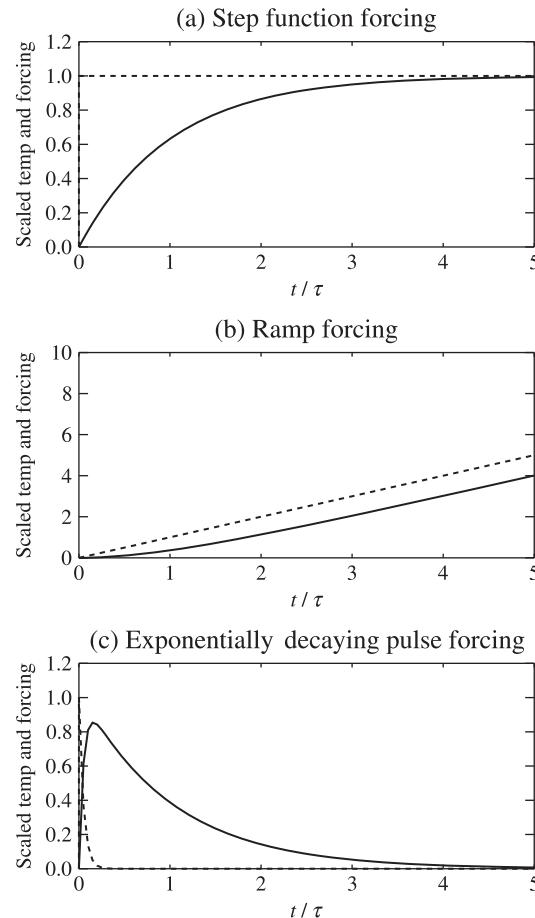


Fig. 8.2 The EBM's temperature response (solid lines) to three different forcings (dashed lines), plotted against non-dimensional time t/τ . (a) Response to step function forcing, given by equation (8.12) with $F_1 > 0$; the scaled temperature is T'/S and the scaled forcing is F/F_1 . (b) Response to ramp forcing, given by equation (8.15) with $\gamma > 0$; the scaled temperature is $T'\alpha/(\gamma\tau)$ and the scaled forcing is $F/(\gamma\tau)$. (c) Response to exponentially decaying pulse forcing, given by equation (8.17) with $F_1 > 0$ and $t_0 = 0.05\tau$; the scaled forcing is F/F_1 , and the scaling of temperature is discussed in Problem 8.3.

increases linearly, like St/τ , for $t \ll \tau$; its initial slope $S/\tau = F_1/C$ by equations (8.9) and (8.14), and hence is independent of the feedback parameter α . However, the feedback comes into play for $t \gtrsim \tau$ and T' then starts to level off, reaching $0.95S$ at about $t = 3\tau$.

Ramp forcing

Here the radiative forcing increases linearly with time, for $t > 0$:

$$F(t) = 0 \quad \text{for } t \leq 0, \quad = \gamma t \quad \text{for } t > 0, \quad (8.15)$$

where γ is constant. This is a good representation of the radiative forcing due to increasing carbon dioxide, since the CO₂ mixing ratio is currently increasing exponentially in time, by about 0.5% per year, and the radiative forcing goes roughly like the logarithm of the CO₂ mixing ratio: see [equation \(8.24\)](#). In this case the solution is a little more difficult to obtain, but turns out to be

$$T'(t) = \frac{\gamma\tau}{\alpha} \left(\frac{t}{\tau} - 1 + e^{-t/\tau} \right) \quad \text{for } t > 0. \quad (8.16)$$

[Figure 8.2\(b\)](#) shows the time-development of the forcing and the solution. It can be checked from [equation \(8.16\)](#) that T' initially (for $t \lesssim \tau$) grows quadratically with time, but later (for $t \gtrsim \tau$) the growth becomes linear with time: this behaviour is clear from the diagram.

Exponentially decaying pulse forcing

Here we take

$$\begin{aligned} F(t) &= 0 \quad \text{for } t \leq 0, \\ &= F_1 e^{-t/t_0} \quad \text{for } t > 0, \end{aligned} \quad (8.17)$$

where F_1 and t_0 are constants. With $F_1 < 0$ this might be a representation of the forcing due to the injection of aerosols into the stratosphere from a massive volcanic eruption. [Figure 8.2\(c\)](#) shows the time-development of the forcing and the solution for the case $t_0 = 0.05\tau$, corresponding for example to values of $t_0 = 1.5$ years and $\tau = 30$ years. Note that the response is spread over a time scale of order τ , which in this case is much longer than the time scale t_0 of the pulse. See also [Problem 8.3](#).

Sinusoidal forcing

Here we take

$$F(t) = F_2 \sin(\omega t) \quad (8.18)$$

for all t , where F_2 is a constant, and seek a purely oscillatory response. This could approximately represent forcing by the 11-year solar cycle, for example. See [Problem 8.4](#).

Note that the assumption $\alpha > 0$ implies that all of these solutions are stable, in the sense that the temperature response cannot grow faster than the forcing. It can easily be checked that if $\alpha < 0$ the solutions will grow exponentially with time, leading to unstable climate change in the EBM.

8.3.1 Response to forcing that initially increases linearly with time and then becomes constant

To introduce some of the issues that arise in detailed climate-change modelling, we now consider the response to ‘ramp-flat’ forcing, which starts as a linear growth with time and then becomes steady at a specified time t_1 , say: this could correspond to increasing CO₂

forcing for time t_1 followed by a stabilised CO₂ amount. The solution is equivalent to a positive ramp followed by an equal and opposite negative ramp, and can be written

$$\begin{aligned} T'(t) &= \frac{\gamma\tau}{\alpha} \left(\frac{t}{\tau} - 1 + e^{-t/\tau} \right) \quad \text{for } 0 < t \leq t_1, \\ &= \frac{\gamma\tau}{\alpha} \left(\frac{t_1}{\tau} - e^{-(t-t_1)/\tau} + e^{-t/\tau} \right) \quad \text{for } t \geq t_1. \end{aligned} \quad (8.19)$$

Equation (8.19) implies that as $t \rightarrow \infty$ (or more precisely for $t - t_1 \gg \tau$), the temperature deviation tends to an ‘equilibrium’ value:

$$T'(t) \rightarrow T_{\text{eqm}} \equiv \frac{\gamma t_1}{\alpha}. \quad (8.20)$$

This equilibrium temperature perturbation equals the forcing at time t_1 , $F(t_1) = \gamma t_1$, divided by α : this is analogous to the definition of climate sensitivity in equation (8.14). Moreover T' approaches T_{eqm} from below, so T_{eqm} is an upper bound on the temperature in this model.

Figure 8.3(a) shows an example of the forcing as a function of time, with $t_1 = 70$ years and $\gamma t_1 = 3.7 \text{ W m}^{-2}$. Figure 8.3(b) shows the corresponding temperature response for three different values of τ . Note that when τ is much less than t_1 (solid curve), the response

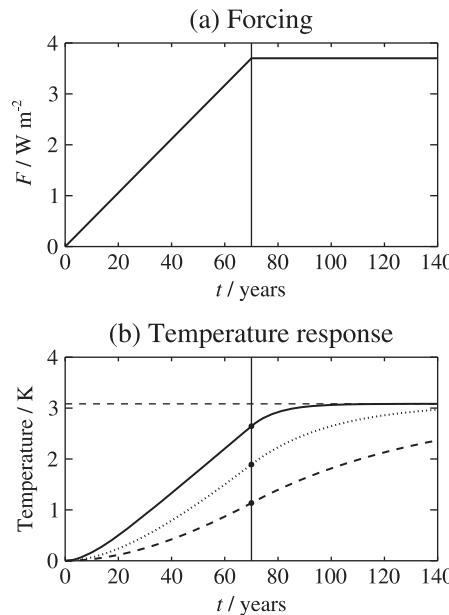


Fig. 8.3

Time dependence of the temperature response in the EBM to ‘ramp-flat’ forcing, with the feedback parameter $\alpha = 1.2 \text{ W m}^{-2} \text{ K}^{-1}$. (a) The forcing as a function of time, with $t_1 = 70$ years and $\gamma t_1 = 3.7 \text{ W m}^{-2}$. (b) The temperature response for $\tau = 10$ years (solid), 30 years (dotted) and 70 years (dashed); these correspond to three different values of C , by equation (8.9), since α is specified. The thin horizontal dashed line shows T_{eqm} , the thin vertical line shows time t_1 and the black circles show $T'(t_1)$.

equilibrates fairly soon after time t_1 ; however, when $\tau = t_1$ (dashed curve), the approach to equilibrium is very much slower.

An important goal for climate-change modelling is to forecast the equilibrium temperature that would result from stabilisation of CO₂ emissions, given information up to some finite time, for example the stabilisation time $t = t_1$. This is straightforward in principle for the ramp-flat model, provided that the parameters are known. However, a practical difficulty for climate modelling is that (as noted in Sections 8.2 and 8.3) the values of the parameters α and C for the Earth's climate system cannot be specified with confidence; hence the FRT τ is not well constrained. Any theory has to allow for uncertainty in τ , and this will lead to uncertainty in the ratio $T_{\text{eqm}}/T'(t_1)$ and hence in the forecast equilibrium temperature based on the value at time $t = t_1$. Figure 8.3(b) illustrates how this ratio varies strongly as τ ranges from $t_1/7$ to t_1 ; see also Problem 8.6.

8.4 Climate feedbacks

We now look at climate feedbacks in more detail, taking account of the fact that the temperature dependence of $Q = F^\downarrow - F^\uparrow$ comes about partly through the dependence of Q on a number of variables V_i , such as water vapour concentration, clouds, albedo and tropospheric lapse rate, each of which depends on temperature. The quantity Q can also depend on a further set of variables U_j , such as carbon dioxide concentration, that are independent of temperature. We therefore replace $Q = Q(T, U)$ by

$$Q = Q(T, V_1(T), V_2(T), \dots, V_N(T), U_1, U_2, \dots, U_M). \quad (8.21)$$

The definition (8.6) of climate feedback is replaced by

$$\alpha = -\frac{\partial Q}{\partial T} - \sum_{i=1}^N \frac{\partial Q}{\partial V_i} \frac{dV_i}{dT},$$

where the term $\partial Q/\partial T$ refers to the partial derivative of Q with respect to the first T argument in equation (8.21), keeping all the V_i and U_j constant, and all derivatives are evaluated at the steady state values $(T, U_j) = (T_{\text{ss}}, U_{j\text{ss}})$. The feedback parameter α may be separated into its constituent parts,

$$\alpha = \sum_{i=0}^N \alpha_i \quad \text{where} \quad \alpha_0 \equiv -\frac{\partial Q}{\partial T} \quad \text{and} \quad \alpha_i \equiv -\frac{\partial Q}{\partial V_i} \frac{dV_i}{dT} \quad \text{for } i \geq 1.$$

It is found that α_0 is generally larger in magnitude than the other α_i . However, some of the latter may be non-negligible compared with α_0 . Moreover, some of these – associated for example with water vapour (see Section 8.4.2) and clouds – may have the opposite sign to α_0 , and hence significantly reduce the total α .

With the form of Q given in equation (8.21) we can also define the radiative forcing associated with a perturbation U'_j in the variable U_j as

$$F_j = \frac{\partial Q}{\partial U_j} U'_j, \quad (8.22)$$

by analogy with [equation \(8.7\)](#).

8.4.1 Black-body feedback

The simplest example of feedback occurs when the climate system acts as a black body at unperturbed temperature T_{ss} , and modifying effects due to the atmosphere are neglected. We assume $F^\downarrow = F_0 \approx 240 \text{ W m}^{-2}$, independent of temperature; the Stefan–Boltzmann law [\(1.1\)](#) gives $F^\uparrow = \sigma T_{ss}^4$, so

$$Q = F^\downarrow - F^\uparrow = F_0 - \sigma T_{ss}^4 = 0 \quad (8.23)$$

in the steady state. Hence T_{ss} equals the emission temperature $T_e \approx 255 \text{ K}$, as in [Section 1.3.1](#). Taking $\alpha_0 = -\partial Q/\partial T_{ss}$ gives

$$\alpha_0 = 4\sigma T_{ss}^3 = \alpha_{BB} \equiv 4\sigma T_e^3 \approx 3.8 \text{ W m}^{-2} \text{ K}^{-1};$$

we call α_{BB} the **black-body feedback**. The fact that $\alpha_{BB} > 0$ expresses the physical result that the power radiated by a black body increases with its temperature.

In the presence of an atmosphere the situation is more complicated. [Equation \(8.23\)](#) can be replaced by

$$Q = F^\downarrow - F^\uparrow = F_0 - \sigma T_e^4 = 0$$

and if we have a functional relation between T_e and T_{ss} we can write

$$\alpha_0 = -\frac{\partial Q}{\partial T_{ss}} = 4\sigma T_e^3 \frac{dT_e}{dT_{ss}}.$$

A simple model would be to take $T_e^4 = \varepsilon T_{ss}^4$, where $\varepsilon < 1$ is an equivalent emittance for the atmosphere, assumed independent of temperature. In this case

$$\alpha_0 = 4\sigma T_e^3 \varepsilon^{1/4} = \varepsilon^{1/4} \alpha_{BB}.$$

An example is the radiative greenhouse model of [Section 1.3.2](#), for which it is found from [equation \(1.5\)](#) that $\varepsilon = (1 + \mathcal{T}_{lw})/(1 + \mathcal{T}_{sw})$. However, it should be borne in mind that there may not be any simple relationship between the global-mean values of T_e and T_{ss} in the real atmosphere. Detailed general circulation models of the Earth's climate can be used to estimate α_0 by applying height-independent temperature perturbations in the troposphere. This gives a value of about $3.2 \text{ W m}^{-2} \text{ K}^{-1}$ corresponding to $\varepsilon \approx 0.5$.

8.4.2 Water vapour feedback

Here $Q = Q(T, V(T))$, where $V(T)$ is some measure of the water vapour content of the atmosphere. In our simple zero-dimensional model we could, for example, take V to be the

saturation mixing ratio at the mean surface pressure p_* ,

$$V(T) = \mu_s(T, p_*) \equiv \mu_*(T).$$

This dependence of $Q = F^\downarrow - F^\uparrow$ on water vapour comes mainly from the OLR F^\uparrow , although F^\downarrow may also depend on water, through its dependence on water-vapour absorption in the stratosphere and on the albedo, which in turn depends on clouds and ice.

The corresponding feedback parameter is

$$\alpha_{\text{wv}} = -\frac{\partial Q}{\partial \mu_*} \frac{d\mu_*}{dT},$$

and from the Clausius–Clapeyron equation (2.37), using $\mu_*(T) = \epsilon e_s(T)/p_*$ from equation (2.40), we have

$$\frac{d\mu_*}{dT} = \frac{\mu_* L}{R_v T^2},$$

which is clearly positive. Moreover, $\partial Q/\partial \mu_*$ is also generally positive, since the OLR F^\uparrow decreases, and hence $Q = F^\downarrow - F^\uparrow$ increases, with increasing amount of water vapour; therefore α_{wv} , unlike α_{BB} , is negative. This implies that as the Earth warms (for example as a result of increasing CO₂), increasing water vapour amounts contribute a decrease in OLR and so tend to warm the climate further.² This is an example of **positive feedback**.

There is some ambiguity in the climate-change literature over the definitions of ‘positive’ and ‘negative’ feedback, and even in the definition of ‘feedback’ itself. The convention we have used here is to define a non-black-body feedback α_i as ‘positive’ if $\alpha_i/\alpha_{\text{BB}} < 0$ and as ‘negative’ if $\alpha_i/\alpha_{\text{BB}} > 0$. However, an alternative definition is based on the sign of the *total* feedback parameter α ; as noted at the end of Section 8.3, our simple EBM is stable if $\alpha > 0$ and unstable if $\alpha < 0$.

Detailed calculations show that $-\alpha_{\text{wv}}/\alpha_0 \approx 0.4$ under present conditions. While this implies that water vapour contributes a significant positive feedback, it is not sufficient to make the sum $\alpha \approx \alpha_0 + \alpha_{\text{wv}}$ negative; hence the climate in this model is stable. If $-\alpha_{\text{wv}}/\alpha_0$ were to exceed 1, an unstable ‘runaway climate’ could develop; however, such a possibility seems to require conditions that are far warmer than those encountered in the present-day global-mean terrestrial climate. It has, however, been speculated that this runaway greenhouse effect may have occurred on the planet Venus.

8.4.3 Other feedback processes

The net inward heat flux Q can depend on temperature in many other ways, and for each of these we can define a climate feedback. Examples include feedback processes associated with clouds, lapse rate, ice-albedo (see Problem 8.7), the carbon cycle, ocean circulation, etc. Some of these are highly complex and are poorly understood at present: much current

² However, it should be noted that the calculation used here is grossly over-simplified: the real atmosphere is not at a uniform temperature, and we should really be considering an ensemble of air-parcel trajectories, with each parcel retaining its mixing ratio until saturated. See Pierrehumbert *et al.* (2007).

research is devoted to trying to understand and quantify the effects of these feedbacks on climate change.

8.5 The radiative forcing due to an increase in carbon dioxide

Detailed calculations show that the radiative forcing due to a change of CO₂ is approximately proportional to the logarithm of the fractional change of CO₂ mixing ratio: if the CO₂ amount increases by a factor β , the radiative forcing is given to a good approximation by

$$F_{\beta \times \text{CO}_2} = 5.3 \ln \beta \quad \text{W m}^{-2}. \quad (8.24)$$

In this section we give a simple, approximate, demonstration of the main physical principles that underlie this logarithmic dependence. We find that the spectral variation of the transmittance of CO₂ is an essential feature: a ‘grey gas’ model would fail completely.

The simplest relevant model takes the ground to be a black body at a ‘warm’ temperature T_g , with Planck function $B_{vg} \equiv B_\nu(T_g)$ (see [equation \(3.1\)](#)) at frequency ν , and represents the troposphere by a thin slab at a uniform ‘cool’ temperature $T_t < T_g$, with Planck function $B_{vt} < B_{vg}$ and spectral transmittance \mathcal{T}_ν . The upward spectral irradiance at frequency ν above the troposphere is given by emission from the ground (with emittance 1), decreased by the tropospheric transmittance, plus emission from the troposphere (with emittance $1 - \mathcal{T}_\nu$):

$$F_\nu^\uparrow = \pi \mathcal{T}_\nu B_{vg} + \pi(1 - \mathcal{T}_\nu)B_{vt}. \quad (8.25)$$

(Note the analogy with the first line of [equation \(7.3\)](#) for the radiance received along a single path through the whole depth of the atmosphere.) We refer to the upward spectral irradiance as the **spectral outgoing long-wave radiation**, or SOLR for short.

The minimum possible value of \mathcal{T}_ν is 0, corresponding to a totally opaque troposphere; in this case F_ν^\uparrow equals the black-body irradiance πB_{vt} from the (top of) the troposphere. The maximum possible value of \mathcal{T}_ν is 1, corresponding to a totally transparent troposphere, when F_ν^\uparrow equals the surface black-body irradiance πB_{vg} . For $0 < \mathcal{T}_\nu < 1$, [equation \(8.25\)](#) shows that F_ν^\uparrow lies between these two extreme values.

For a single absorbing gas the tropospheric transmittance \mathcal{T}_ν can be approximated by [equation \(3.23\)](#) modified by the diffusivity factor (see [equation \(3.16\)](#)):

$$\mathcal{T}_\nu = \exp \left(-1.66 \int k_\nu(z) \rho_a(z) dz \right). \quad (8.26)$$

Here the integral is taken over the depth of the troposphere, k_ν is the spectral absorption coefficient and ρ_a is the absorber density.

We wish to find the radiative forcing due to a change in density of this absorber. Neglecting any short-wave absorption, this radiative forcing is minus the change of SOLR just above the troposphere, integrated over all frequencies. Let us suppose that the absorber density increases by a factor $\beta > 1$ at each level z , from $\rho_a(z)$ to $\beta \rho_a(z)$. [Equation \(8.26\)](#)

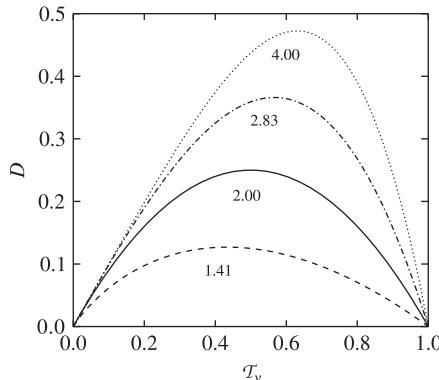


Fig. 8.4

Plots of $D(\mathcal{T}_v)$ against \mathcal{T}_v for $\beta = \sqrt{2}, 2, 2\sqrt{2}$ and 4.

shows that, under this change, the tropospheric transmittance is raised to the power β , i.e. it goes from \mathcal{T}_v to \mathcal{T}_v^β . In accordance with the definition of radiative forcing (see Section 8.1), we assume that the temperatures of the ground and troposphere do not change under this absorber change; the resulting change ΔF_v^\uparrow in SOLR above the troposphere is therefore

$$\Delta F_v^\uparrow = \pi (B_{vg} - B_{vt}) \Delta \mathcal{T}_v. \quad (8.27)$$

Here we have defined the change of tropospheric transmittance

$$\Delta \mathcal{T}_v \equiv \mathcal{T}_v^\beta - \mathcal{T}_v \equiv -D(\mathcal{T}_v, \beta), \text{ say.} \quad (8.28)$$

For $\beta > 1$ and $0 < \mathcal{T}_v < 1$ this change of transmittance is negative, and D is positive: hence equation (8.27) implies that the upward irradiance at the tropopause decreases under the increase of absorber. Alternatively, in terms of equation (8.25), as the transmittance decreases less SOLR comes from the warm surface and more from the cool troposphere, leading to an overall decrease of the SOLR. However, $D = 0$ when $\mathcal{T}_v = 0$ or 1, reflecting the fact that the transmittance cannot change at frequencies where the troposphere is already totally opaque or totally transparent.

Figure 8.4 shows the decrease in transmittance $D(\mathcal{T}_v, \beta)$ for the full range $0 \leq \mathcal{T}_v \leq 1$, for the four values $\beta = \sqrt{2}, 2, 2\sqrt{2}$ and 4. Note that D attains a maximum value $D_{\max}(\beta)$ somewhere in mid-range, with D_{\max} increasing with β . It can be shown that, for $1 < \beta \leq 4$, $D_{\max}(\beta)$ is given to a good approximation by

$$D_{\max} = \frac{1}{e} \ln \beta. \quad (8.29)$$

(See Problem 8.8.)

The value of \mathcal{T}_v at each frequency ν , and hence the SOLR and the radiative forcing, for a particular absorbing gas will depend on the amount of absorber and its spectroscopic properties. We take the absorber to be carbon dioxide and concentrate on its main absorption band: Figure 8.5(a) plots \mathcal{T}_v against ν at high resolution and Figure 8.5(b) plots a smoothed version of \mathcal{T}_v (cf. Figure 3.14, which includes a low-resolution plot of the transmittance of a vertical atmospheric column, plotted against wavelength). Focusing for clarity on the

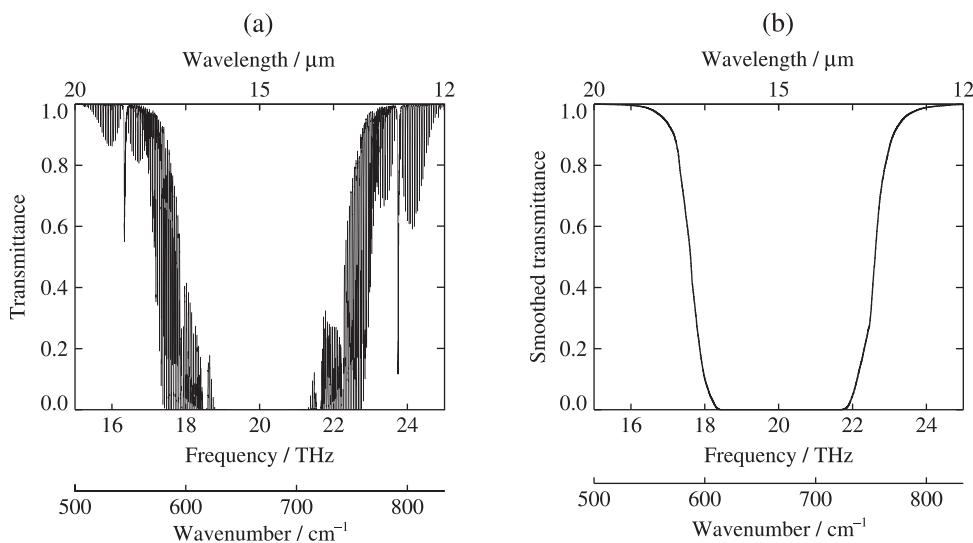


Fig. 8.5

(a) Spectrum of T_ν for the main absorption band of CO_2 , plotted against frequency ν in THz and also against wavenumber $\tilde{\nu}$ in cm^{-1} and wavelength μm , at a resolution of 0.3 GHz (0.01 cm^{-1}), based on data provided by Dr A. Dudhia. (b) A smoothed version of the same spectrum: this is obtained by reordering the original spectrum with respect to frequency, to make it monotonic in frequency on either side of the centre of the band. The range of variation of T_ν and the area under the curve are unchanged by the reordering.

smoothed spectrum, we note that, as ν increases, T_ν drops from 1 to 0 between about 16 and 18 THz and then rises from 0 to 1 between about 22 and 24 THz. The maximum value D_{\max} of $D(T_\nu)$ will be attained at certain frequencies in these regions on the ‘flanks’, or ‘wings’, of the absorption band.

Outside these regions ΔT_ν equals either 1 – 1 or 0 – 0, i.e. it is zero; so by equation (8.27) the change ΔF_ν^\uparrow in SOLR is also zero there. Physically, there can be no change in SOLR at those frequencies where the troposphere is either completely transparent (optically thin) or completely opaque (optically thick): neither of these cases is affected by an increase in absorber density.

Figure 8.6(a) shows the decrease in SOLR for the case $\beta = 2$, with $T_g = 290 \text{ K}$ and $T_t = 225 \text{ K}$ chosen to represent a typical observed mid-latitude upper-tropospheric temperature. This decrease has a peak, with maximum value $\pi(B_{vg} - B_{vt})D_{\max}$, on each flank of the absorption band. Figure 8.6(b) shows the decrease in SOLR divided by $\ln \beta$, for $\beta = 2$ and 4; the fact that the two curves are so close together indicates that the approximate logarithmic dependence of D_{\max} in equation (8.29) carries over to $-\Delta F_\nu^\uparrow$ at each frequency.

Another way of looking at this result is to plot the unperturbed SOLR, and the SOLR under a doubling of CO_2 , against frequency: see Figure 8.7. Also plotted are πB_ν for the surface and the troposphere (which is again given a typical observed upper-tropospheric temperature). As expected from equation (8.25), both SOLR curves lie between the two Planck function curves, rising to meet the surface curve outside the absorption band, and

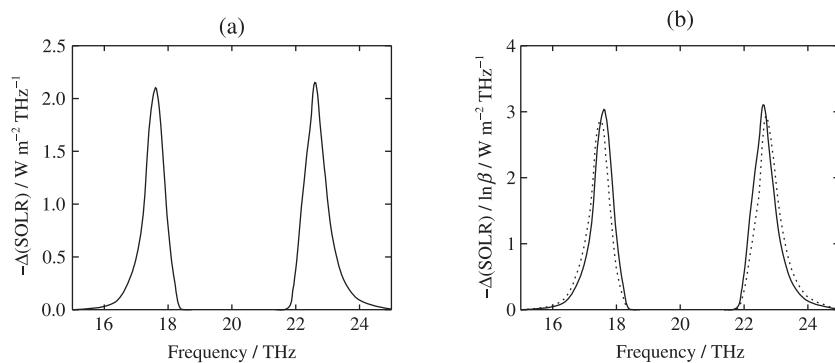


Fig. 8.6 (a) Decrease $-\Delta F_v^\uparrow$ in SOLR given by equation (8.27) for a doubling of CO₂ density ($\beta = 2$), based on the smoothed spectrum on the right of Figure 8.5, with $T_g = 290\text{ K}$ and $T_t = 225\text{ K}$. (b) $-\Delta F_v^\uparrow / \ln \beta$, for $\beta = 2$ (solid) and 4 (dotted).

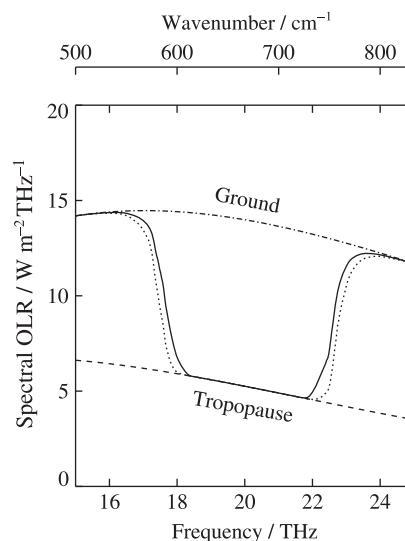


Fig. 8.7 Unperturbed SOLR F_v^\uparrow (solid), and the SOLR under a doubling of CO₂ (dotted), based on the smoothed spectrum on the right of Figure 8.5, plotted against frequency and wavenumber. Also shown are πB_v evaluated at the surface (dot-dashed) and at the tropopause (dashed), with $T_g = 290\text{ K}$ and $T_t = 225\text{ K}$.

falling to meet the tropopause curve inside it. As CO₂ doubles, the SOLR decreases a little at each wavenumber on the flanks of the CO₂ band. The decrease from the unperturbed case is given by the vertical distance between the two SOLR curves, and this is equal to the function plotted in Figure 8.6(a). This decrease leads to a broadening of the ‘bite’ taken out of the surface Planck function curve: see Pierrehumbert (2010).

The final stage in the calculation of the radiative forcing is to integrate the spectral results over frequency, i.e. to calculate the areas under the curves in Figure 8.6(b).³ (See Problem 8.9.) The fact that these areas depend only weakly on β shows that the approximate logarithmic dependence carries over to the radiative forcing F , which therefore takes the form

$$F = - \int_0^\infty \Delta F_v^\uparrow dv \approx A \ln \beta, \quad (8.30)$$

where A is a constant, in agreement with the form of equation (8.24). Note that the logarithmic dependence does not depend on the detailed structure of the spectral lines, but just on the fact that the tropospheric transmittance varies rapidly in frequency between 0 and 1 on each flank of the absorption band, so that the most significant values of the transmittance change (almost proportional to $\ln \beta$) are attained in a narrow range of frequencies there. The integral is dominated by contributions from frequencies in these flank regions.

For comparison with Section 8.2, it is convenient to put $U = \ln \rho_0$, where $\rho_0 = \rho_a(0)$ is the CO₂ density at the ground. The change from ρ_0 to $\beta \rho_0$ then corresponds to the finite change $\Delta U = \ln \beta$, so equation (8.30) gives $F \approx A \Delta U$. This implies that equation (8.7) holds without the ‘small perturbation’ assumption, given $\partial Q/\partial U = A$.

The CO₂ case should be contrasted with that of absorbers such as the chlorofluorocarbons, which are optically thin at all frequencies, so that T_v is always close to 1 (i.e. it is always close to the right-hand end of the range shown in Figure 8.4), and D_{\max} is never attained at any v . In this case we can show (e.g. by putting $T_v = 1 - A_v$, where $A_v \ll 1$) that $D(T_v) \sim (\beta - 1)(1 - T_v)$. Therefore for such absorbers the radiative forcing is linear, not logarithmic, in β : this is confirmed by detailed calculations.

Of course, our simple slab model is an extremely crude representation of the real troposphere. However, it can be extended to a multi-level model in which the temperature follows a specified lapse rate; the argument that we have used is applied at each level within the troposphere, and then a vertical integral performed. Approximate logarithmic dependence of the SOLR at each frequency is found, and essentially the same results follow.

Further reading

A good introduction to the physics of climate change is given by Hartmann (1994). The advanced but readable book by Pierrehumbert (2010) covers the material of this chapter in more detail, and goes much further – including, for example, the physics of climate change in the Earth’s distant past and on other planets. The Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Working Group I (Solomon *et al.* (2007)), gives a wealth of up-to-date information on climate change and the physical processes contributing to it. Houghton (2009) provides an authoritative overview of the current

³ This is almost identical to integrating the unsmoothed functions, since the smoothing just reorders them with respect to frequency. The Planck function is not reordered; however, it only varies slowly with frequency. Note also that other CO₂ absorption bands are present, as shown by Figure 3.14, but these are less important because B_v is much smaller for the frequencies and temperatures involved.

understanding of the science of climate change and its impacts on human society. See Bates (2007) for a thorough discussion of the concept of ‘climate feedback’. The paper by Myhre *et al.* (1998) provides detailed calculations of the radiative forcing due to changes in CO₂ and other greenhouse gases.

Problems

Problem 8.1 Consider a model troposphere in the form of a thin slab at a temperature T_t above a warmer black surface at temperature T_g . (This system need not be in radiative equilibrium, i.e. non-radiative heat transfers could occur.) Suppose that the troposphere contains a greenhouse gas whose absorption coefficient is independent of frequency in the infra-red (this is called a ‘grey’ absorber) and whose infra-red transmittance is \mathcal{T} . Show that the upward irradiance above the troposphere is given by

$$F^\uparrow = \mathcal{T}\sigma T_g^4 + (1 - \mathcal{T})\sigma T_t^4, \quad (8.31)$$

where σ is Boltzmann’s constant.

Assuming that F^\uparrow balances the unreflected incoming solar irradiance $F_0 = 240 \text{ W m}^{-2}$, and given that the tropospheric temperature $T_t = 245 \text{ K}$, find the surface temperature T_g for $\mathcal{T} = 0.1, 0.2$ and 0.3 . Show that the ‘greenhouse effect’ of the gas, defined as $\sigma T_g^4 - F_0$, is proportional to $\mathcal{T}^{-1} - 1$ and evaluate it in each case. Repeat the calculations for $T_t = 225 \text{ K}$. Comment on the implication of your answers.

Problem 8.2 Suppose that the troposphere in the previous problem contains two greenhouse gases that are both grey absorbers and whose infra-red transmittances are \mathcal{T}_1 and \mathcal{T}_2 , both less than 1. Show that their combined transmittance is $\mathcal{T}_1\mathcal{T}_2$. Hence show that the greenhouse effect of the combination of gases is not equal to the sum of their separate greenhouse effects, except (approximately) when both gases are optically thin, i.e. when \mathcal{T}_1 and \mathcal{T}_2 are both close to 1. (Hint for last part: put $\mathcal{T}_1 = 1 - \mathcal{A}_1$, where \mathcal{A}_1 is small, etc.)

Problem 8.3 Find an expression for the response of the EBM (8.8) to the exponentially decaying pulse forcing (8.17). Verify the plot of the time-dependence of T' in Figure 8.2(c), given $\tau = 30$ years and $t_0 = 1.5$ years. What scaling is used for T' in that figure? Find the maximum negative temperature response, given $F_1 = -2 \text{ W m}^{-2}$.

Problem 8.4 Find expressions for the response of the EBM (8.8) to the sinusoidal solar cycle forcing (8.18), ignoring transients, i.e. find T' in the form $a \sin(\omega t - \phi)$, where the amplitude a and phase lag ϕ are to be obtained. (Hint: it may be helpful to use complex exponentials.) What are the values of the amplitude and the phase lag, given $\tau = 30$ years and $F_2 = 0.12 \text{ W m}^{-2}$?

A hypothetical ‘equilibrium response’ to the forcing can be defined as the solution for the case when the time-derivative on the left of equation (8.8) is ignored. Evaluate the ratio

of a to the amplitude of the equilibrium response, given τ as above. Sketch the forcing, the full solution and the equilibrium solution on the same time axis.

Problem 8.5

1. Using the formula given in [equation \(8.24\)](#), evaluate the radiative forcing $F_{2\times\text{CO}_2}$ associated with a doubling of CO_2 .
2. Assuming an increase in CO_2 concentration of 0.5% per year, evaluate the constant $\gamma = dF/dt$ in [equation \(8.15\)](#). If this rate were sustained, how many years would it take for the concentration to double? How long would it take for the volume mixing ratio to increase from a recent value of 385 ppmv (cf. [Table 2.1](#)) to 550 ppmv (approximately double the pre-industrial value)?
3. Assuming that the only feedback process is black-body feedback, evaluate the climate sensitivity under an instantaneous doubling of CO_2 . Given that the climate system has a heat capacity C per unit horizontal area of $1 \text{ GJ K}^{-1} \text{ m}^{-2}$, evaluate the corresponding feedback response time (FRT) τ .
4. The climate sensitivity in a complex climate-forecast model is in fact found to be 3 K under an instantaneous doubling of CO_2 . What value of the feedback parameter α does this correspond to? Assuming the same value of C as above, evaluate the FRT in this case. Comment on the effects of non-black-body feedbacks in this model.
5. Now suppose that errors in the climate-forecast model imply that the value of α obtained above is uncertain to within $\pm 50\%$. What are the corresponding ranges of uncertainty of climate sensitivity and FRT? Comment on the distribution of these ranges about the values established in the previous paragraph.

Problem 8.6 For the ‘ramp-flat’ model of [Section 8.3.1](#), show that the ratio of the equilibrium temperature perturbation to that at time t_1 is a function of $x = t_1/\tau$ only. Show that this function is large for small x and tends to 1 for large x , and give a rough sketch of it. Confirm that the shape of this function agrees with the results plotted in [Figure 8.3](#).

For the three values of τ used in [Figure 8.3](#), find the times at which T' reaches 90% and 99%, respectively, of its equilibrium value T_{eqm} .

Problem 8.7 Consider a simple energy balance model including ice-albedo feedback, in which the Earth’s albedo A varies strongly with the amount of ice cover: represent this by allowing A to vary with temperature as follows:

$$A = 0.6 \quad \text{for } T \leq 265 \text{ K},$$

$$A = 0.3 \quad \text{for } T \geq 275 \text{ K},$$

$$A = 0.45 + 0.03 \times (270 \text{ K} - T) \quad \text{for } 265 \text{ K} \leq T \leq 275 \text{ K}.$$

Hence modify [equation \(8.2\)](#) for F^\downarrow appropriately; assume $F^\uparrow(T) = 0.6\sigma T^4$, where the emissivity factor 0.6 is a simple representation of the greenhouse effect.

1. Give a brief physical explanation of the formula for $A(T)$.
2. Show that this model has equilibrium (steady-state) solutions for a temperature $T_1 < 265$ K ('snowball Earth') and a temperature $T_3 > 275$ K, and evaluate T_1 and T_3 . Draw a sketch of $F^\downarrow(T)$ and $F^\uparrow(T)$ as functions of T . Show from your sketch that there is also an equilibrium temperature T_2 between 265 and 275 K. (Do not evaluate T_2 : it is in fact about 269 K).
3. Find an expression for the feedback parameter $\alpha = -dQ/dT$ for this model, where $Q(T) = F^\downarrow - F^\uparrow$. From your sketch determine the sign of α at each equilibrium temperature, and hence infer whether each equilibrium is stable or unstable. Comment on the concept of 'ice-albedo feedback' in this model.

Problem 8.8 Given the function D defined by equation (8.28), show that its maximum value for each β is

$$D_{\max} = \beta^{-1/(\beta-1)} - \beta^{-\beta/(\beta-1)}.$$

Now show that D_{\max} can be written in terms of $u \equiv \ln \beta$ as

$$D_{\max} = \exp \left\{ \frac{-u}{e^u - 1} \right\} - \exp \left\{ \frac{-u}{1 - e^{-u}} \right\}.$$

Verify that D_{\max} is an odd function of u . By expanding to the first power in u show that

$$D_{\max} \sim \frac{u}{e} = \frac{1}{e} \ln \beta \quad \text{as } u \rightarrow 0, \quad \text{i.e. as } \beta \rightarrow 1. \quad (8.32)$$

Without detailed calculation show that the next term in the expansion must be of order u^3 . Verify that the approximation (8.29) for D_{\max} is about 8% too large at $\beta = 4$.

Problem 8.9 Use Figure 8.6 to obtain a rough estimate of the coefficient A of $\ln \beta$ in equation (8.30) for the radiative forcing for CO₂.

This chapter is a short introduction to the use of models in atmospheric research and forecasting. In [Section 9.1](#) we explain how a hierarchy of models – simple, intermediate and complex – can be used for gaining understanding of atmospheric behaviour and interpreting atmospheric data. In [Section 9.2](#) we give brief details of the numerical methods used in the more complex theoretical models, while in [Section 9.3](#) we outline the use of these models for forecasting and other purposes. In [Section 9.4](#) we describe an example of a class of laboratory models of the atmosphere. Finally, in [Section 9.5](#), we give some examples of atmospheric phenomena that arise from interactions between basic physical processes and that can be elucidated only with the aid of models of intermediate complexity.

9.1 The hierarchy of models

The basic philosophy of atmospheric modelling was outlined in [Section 1.2](#). It was mentioned there that a hierarchy of models, from simple to complex, must be used for understanding and predicting atmospheric behaviour; this hierarchy is illustrated in [Figure 9.1](#). The simple models ('back-of-the-envelope' or 'toy' models) involve a minimum number of physical components and are described by straightforward mathematical equations that can usually be solved analytically. These models provide basic physical intuition: most of the models considered earlier in this book are of this type. The intermediate models involve a small number of physical components but usually require a computer for solution of the mathematical equations. Simple and intermediate models can provide valuable conceptual pictures of atmospheric processes, but do not usually give accurate simulations of actual atmospheric behaviour. The complex models (often called **general circulation models** or GCMs) contain mathematical representations of a large number of physical processes. These are the models that are used when accurate simulations are required.

Each type of model is motivated by atmospheric observations and each may help us understand some aspect of atmospheric behaviour. The physical intuition provided by the simplest models helps us interpret the intermediate models and the intermediate models help us interpret the complex models. When a complex model fails to produce a reasonable simulation of observed atmospheric behaviour, it may be necessary to use an intermediate model (perhaps a 'stripped-down' version of the complex model) to find the cause.

The types of physical process that can be included in atmospheric models fall into three main categories, each of which has been studied in earlier chapters.

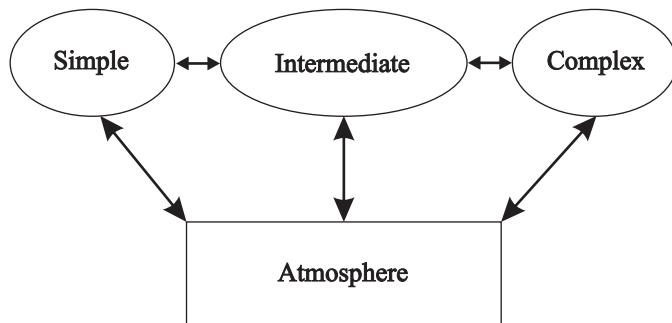


Fig. 9.1

Illustrating the use of a hierarchy of atmospheric models. Each type of model is motivated by, and provides some insight into, atmospheric behaviour. The simpler models help us interpret the more complex models. If a more complex model fails, a simpler model may help us to locate the cause.

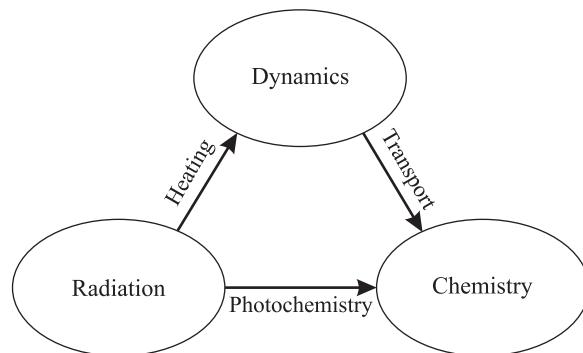


Fig. 9.2

Illustrating some of the interactions among dynamics, radiation and chemistry in the atmosphere; see the text for details.

- **Dynamical processes**, including those parts of thermodynamics associated with the First Law. These were introduced in [Chapters 4 and 5](#).
- **Radiation**: this was introduced in [Chapter 3](#), and some of its consequences for climate change were explored in [Chapter 8](#).
- **Chemistry**: some aspects of this were introduced in [Chapter 6](#).

These three categories interact in complex ways in the atmosphere, as is partly illustrated in [Figure 9.2](#). Transport by dynamical processes carries chemicals from one part of the atmosphere to another, for example vertically from ground level to the stratosphere, or horizontally from the industrialised Northern Hemisphere to Antarctica; see [Section 6.6](#). The heating due to absorption of solar ultra-violet radiation by ozone may drive dynamical processes. Solar radiation may be energetic enough to disrupt chemical bonds, leading to photochemical reactions. In general, models incorporating even a few of these interactions will be too complicated for simple analysis and so will be of ‘intermediate’ or ‘complex’ type.

9.2 Numerical methods

As mentioned in [Section 9.1](#), intermediate and complex models require us to solve the mathematical equations describing the physics by means of computers. This is not an easy task and may require extremely careful and efficient computer programming and huge computer codes. In this section we describe some of the basic principles involved.

The equations describing the dynamics are generally partial differential equations, involving differentiation in space and time. Consider for example the x -momentum equation; under the approximations discussed in [Section 4.7.2](#) this can be written

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} - w \frac{\partial u}{\partial z} + fv - \frac{1}{\rho} \frac{\partial p}{\partial x} + F^{(x)}, \quad (9.1)$$

by [equations \(4.23a\)](#) and [\(4.22\)](#). The physical meanings of the symbols are given in [Section 4.7.2](#), but are not important for the present discussion. The partial time derivative of the eastward velocity u on the left-hand side of [equation \(9.1\)](#) can be approximated by a **finite difference**: from the Taylor expansion

$$u(x, t + \Delta t) = u(x, t) + \Delta t \frac{\partial u(x, t)}{\partial t} + O(\Delta t^2), \quad (9.2)$$

where Δt is assumed small, we have

$$\frac{\partial u(x, t)}{\partial t} \approx \frac{u(x, t + \Delta t) - u(x, t)}{\Delta t}.$$

Now suppose that all the quantities on the right-hand side of [equation \(9.1\)](#) are known at time t ; then this equation gives $\partial u / \partial t$ at time t and substitution into [equation \(9.2\)](#) gives u at time $t + \Delta t$, with an error of order Δt^2 . A ‘time-step’ (in effect a forecast for a tiny time interval) has therefore been performed. In principle, repeated use of this process allows longer forecasts to be made.

The spatial derivatives on the right-hand side of [equation \(9.1\)](#) must also be calculated; one way to do this is again by finite differences. The atmosphere is taken to be spanned by a three-dimensional grid or lattice of points, with spacings Δx , Δy and Δz , say, in the eastward, northward and vertical directions, respectively.¹ Then, for example, a simple ‘centred difference’ approximation to the pressure gradient $\partial p / \partial x$ is

$$\frac{\partial p(x, t)}{\partial x} \approx \frac{p(x + \Delta x, t) - p(x - \Delta x, t)}{2 \Delta x},$$

and other spatial derivatives can be represented in a similar way, thus allowing computation of the terms on the right-hand side of [equation \(9.1\)](#). For current finite-difference models used for numerical weather prediction (NWP), the horizontal grid spacing is typically about 0.6° longitude by 0.4° latitude (corresponding to a grid scale of roughly 40 km in

¹ In spherical geometry special precautions must be taken at the poles. Note also that in complex models pressure coordinates (see [Section 4.9](#)) or other vertical coordinates are usually used instead of the height z .

midlatitudes), with up to 50 vertical levels spaced at 50–100 hPa in pressure (and less than 50 hPa near the ground and the tropopause) and a total of about 10^7 grid-boxes.

An alternative is to calculate horizontal derivatives by **spectral methods**, that is, the expansion of physical quantities in series of orthogonal functions whose derivatives are known explicitly. Consider for example an x derivative again and focus for simplicity on an f -plane (see [Section 4.7.3](#)) that is periodic in the x direction, with a period L that represents the length of the latitude circle. A suitable expansion of $u(x, t)$ is then a Fourier series of the form

$$u(x, t) = \bar{u}(t) + \sum_{n=1}^N \left[a_n(t) \cos\left(\frac{2n\pi x}{L}\right) + b_n(t) \sin\left(\frac{2n\pi x}{L}\right) \right],$$

where N is called the **truncation limit**; the x derivative of u is

$$\frac{\partial u(x, t)}{\partial x} = \sum_{n=1}^N \left[\frac{2n\pi b_n(t)}{L} \cos\left(\frac{2n\pi x}{L}\right) - \frac{2n\pi a_n(t)}{L} \sin\left(\frac{2n\pi x}{L}\right) \right].$$

In spherical coordinates it is convenient to expand in spherical harmonics $\exp(im\lambda)P_n^m(\cos\phi)$, where λ is longitude, ϕ is latitude, P_n^m is an associated Legendre polynomial and $|m| \leq n$. Modern spectral NWP models include harmonics with values of n up to about 800 (again corresponding to a smallest resolved scale of about 40 km); finite differences are still usually used in the vertical. Since the equations of motion such as [equation \(9.1\)](#) are non-linear, it is necessary to re-expand products of pairs of terms, each represented by its own series, into single series. Fast numerical algorithms have been developed for this purpose.

An ever-present danger in numerical modelling is the possibility of numerical instabilities, that is, instabilities of the finite-differenced equations that are not present in the real physical system being modelled. If they are left unchecked, these instabilities may completely swamp the numerical solution; special care must be taken to avoid or inhibit them.

Numerical problems in the modelling of radiation are of a different type. Here the most important quantity to be calculated is the radiative heating rate, which is known in principle from the solution to the radiative-transfer [equation \(3.10\)](#). However, the calculation must be done for several absorbing gases, over a wide variety of wavelengths or frequencies, and for atmospheric paths at different zenith angles. This would involve extensive computations if it were carried out in full, so some of these calculations are often simplified. For example, rather than attempting to integrate in frequency over every known spectral line (the ‘line-by-line’ approach) one may work with a small number of band transmittances (see [equation \(3.25\)](#)), averaged over spectral bands containing many lines. The presence of clouds, which scatter and absorb the radiation, is a further complication.

If chemistry is to be included in a model, then coupled ordinary differential equations representing the chemical reactions (see [Sections 6.2](#) and [6.4](#)) will be needed. Although these appear less complicated than the partial differential equations of dynamics, there may

still be difficulties in solving them numerically. For example, there may be a large number of chemicals involved and hence a large number of equations in the set. Less obviously, the possibility of both fast and slow reactions implies a disparity of natural time scales in the differential equations, which are then said to be ‘stiff’ and require careful numerical treatment.

9.3 Uses of complex numerical models

Over the past few decades the need for reliable weather forecasts has provided a strong impetus for the development of complex atmospheric models, at least for the troposphere and lower stratosphere. Numerical weather prediction models, for predicting a few days to perhaps two weeks ahead, must include representations of large-scale dynamical processes, small-scale frictional processes in the atmospheric boundary layer, small-scale drag due to gravity waves, short-wave and long-wave radiative transfer, water-vapour transport, the effects of clouds, precipitation and the transfer of heat, momentum and moisture between the surface and the atmosphere. Some of these (such as cloud microphysical processes and drag due to small-scale gravity waves) take place on length scales below the grid size and are therefore not explicitly represented in the models. Such processes must be **parameterised**, that is, expressed in terms of quantities that are explicitly represented. Such parameterisations often involve empirical formulae that do not have rigorous physical justification and may involve disposable constants whose values are poorly known. Much of the art of NWP consists in selecting these constants in such a way as to optimise the resulting forecasts.

Forecasts are made by integrating the equations of motion forwards in time, starting from an observed initial state. Owing to measurement error and the inevitable sparsity of data in some regions, the initial state may not be known with great accuracy. This can be a major source of error in forecasts. In common with other chaotic processes, the evolution of the atmosphere can be sensitive to initial conditions. One way to estimate the reliability of the forecast model on any given occasion is to perform many forecasts, starting from slightly different initial conditions. If all members of the ‘ensemble’ of forecasts are similar, then a high degree of reliability can be placed on a representative forecast; however, if the ensemble contains widely differing forecasts, less weight can be placed on any of them.

Complex models are also used for climate forecasting, for years or decades into the future. Here the focus is on the long-term behaviour of the atmosphere, perhaps averaged over seasons, rather than on the detailed day-to-day evolution. On these longer time scales, some physical processes that can be neglected in NWP must be included: for example, a representation of oceanic heat transfer is needed. Much work is being done, for example, to estimate the global and regional responses of the climate system to various possible scenarios for the increase of greenhouse gases over the coming decades. Climate-prediction models of this kind involve very heavy usage of computer resources, not only because they must be run for many simulated years but also because of the variety of physical processes

included and hence the complexity of the mathematical equations to be solved. Moreover, great care must be devoted to examining and interpreting the huge quantity of data produced by these models. However, computer power is now sufficient to allow large ensembles of climate forecasts to be run: in this respect ‘distributed computing’, using the personal computers of a large number of volunteers from the general public, is proving to be a valuable approach.

Complex models of the NWP type are being used increasingly for **data assimilation**. This is a procedure by which observed data (which may be incomplete, or inconsistent because of measurement error) are fed into the model over a period of time. The model can be regarded as an accurate representation of the known laws of physics and dynamics; it adjusts the observations, within their expected error bounds, in such a way as to force them to be closely consistent with these laws. The resulting ‘assimilated’ dataset thus represents a dynamically and physically consistent interpolation of the observed evolution with time of the atmosphere over the observation period. Data assimilation is also used for deriving accurate initial conditions for weather forecasting.

Complex models are heavily used for research purposes, to help understand atmospheric behaviour. In this mode they fit particularly well into the hierarchy described in [Section 9.1](#). A set of controlled experiments may be designed to investigate a particular phenomenon – for example the development of the Antarctic ozone hole in a given year. Simulations of the phenomenon are performed, using various initial conditions, representations of the chemical reactions and so on. Comparison of the resulting simulations with observed data may help to identify which aspects of the initial conditions are most important for the phenomenon or which chemical schemes are closer to reality. Intermediate and simple models, for interpretation of the successful simulations and for investigation of the reasons why others are unsuccessful, may also play crucial roles in such an investigation.

A further use for complex models is in the design of large measurement programmes, by providing representative datasets on which observation strategies and data-analysis schemes can be refined, in advance of the collection of the real data.

9.4 Laboratory models

A quite different approach from those described above is to use a laboratory apparatus to model atmospheric phenomena. This has the advantage of providing a real physical analogue, free from numerical errors; it may also be subjected to detailed and systematic measurement in a way that may not be possible with the atmosphere. On the other hand, great care must be taken to minimise the effects of physical properties of the laboratory system that are irrelevant to the atmosphere.

A good example of a laboratory system that has been used extensively for modelling atmospheric processes is the **rotating annulus**, which was developed by R. Hide in the 1950s and has been used by researchers over many years. The annulus consists of a ring-shaped container of fluid, such as water, which is, for example, heated at the outer wall by

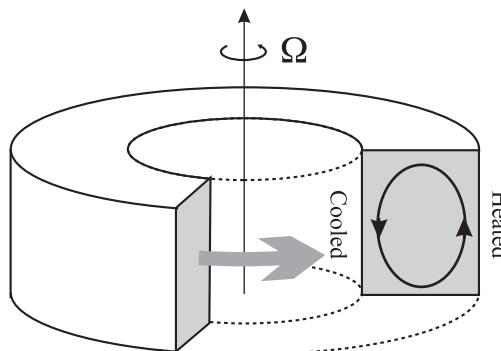


Fig. 9.3

A schematic illustration of the rotating annulus, heated at the outer wall and cooled at the inner wall. A 'cut-away' view is shown, with an indication of the azimuthal, vertical and radial motion found at slow rates of rotation.

maintaining it at temperature T_o and cooled at the inner wall by maintaining it at temperature $T_i < T_o$. The annulus is placed on a rotating turntable and rotated at a constant angular speed Ω about its axis; see Figure 9.3. The fluid in the annulus mimics the atmosphere, the rotation mimics the rotation of the Earth, the heating at the outer radius mimics the heating of the atmosphere in the tropics and the cooling at the inner radius mimics the cooling at high latitudes.

The fluid flow in the annulus takes a number of forms, depending on the speed of rotation and the thermal contrast between the inner and outer walls. At slow rates of rotation, the flow is basically in the azimuthal direction (corresponding to eastward flow in the atmosphere), with a superimposed vertical and radial motion: rising on the heated outer wall, inward flow along the top, descent on the inner wall and outward flow along the base (roughly analogous to the observed mean meridional 'Hadley circulation' in the low-latitude atmosphere). This flow has no azimuthal variation.

At faster rates of rotation the flow is no longer azimuthally symmetric, but rather forms a number of wave-like patterns that drift in azimuth (see Figure 9.4); the number of wavelengths around the annulus depends on the speed of rotation. These waves are related to the baroclinic instabilities discussed in Section 5.7. At certain rates of rotation the number of wavelengths or the amplitude of the waves may pulsate slowly: this pulsation is called **vacillation**. At high rates of rotation, the wave-like flow breaks down to a rather chaotic state reminiscent of some disturbed atmospheric flows. It turns out that an analogue of the β -plane (see Section 4.7.3) can be constructed by sloping the top and bottom walls of the annulus, which allows Rossby waves to be simulated. Extensive observational and theoretical analysis has been carried out on this type of laboratory system, and similar theoretical analysis has been carried out on numerical models of the annulus flow. Despite there being some obvious differences from the atmosphere (for example the much greater role of viscosity and the presence of side-walls in the laboratory analogue) the rotating annulus has provided much insight into atmospheric (and also oceanic) behaviour.

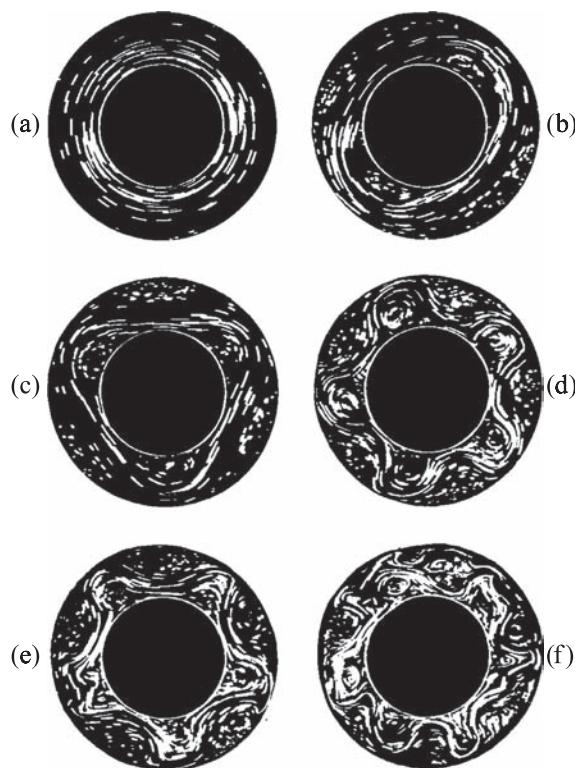


Fig. 9.4

Flow regimes in a rotating annulus in which velocities near the upper surface are made visible by 'streak photography', i.e. by illuminating suspended particles with a flat light beam and taking a time exposure with a camera mounted on the rotating apparatus. The rotation rate Ω increases from frame (a) to frame (f). Note the azimuthally symmetric flow in (a), the increasing number of wavelengths in the azimuthal direction in (b) to (e) and the breakdown to irregular flow in (f). From Hide and Mason (1975), courtesy of Taylor & Francis Ltd, <http://www.informaworld.com>.

9.5 Final remarks

As noted in Section 9.1, most of the models considered in this book have been of the simplest type; we have generally concentrated on single atmospheric processes in isolation, ignoring interactions with other processes. This approach is essential for developing a basic physical intuition for the atmosphere. However, the number of atmospheric phenomena that may be described fully in terms of one type of process alone is limited, so for most purposes interactions between processes must be invoked if a reasonably complete physical description is to be provided. This usually means that models of at least intermediate complexity must be used. In this concluding section we give some examples of such physical descriptions that have been developed with the help of intermediate models.

9.5.1 The height of the tropopause

A climatological zonal-mean temperature field for January was given in Figure 1.5 and some physical processes that help determine this field were outlined in Section 1.4.1. It was noted there that even a basic understanding of the processes determining the position of the tropopause – the interface between the troposphere and stratosphere – involves the interaction between radiation and dynamics. Except in the winter polar regions and in the tropics, radiation is the primary process determining the temperature structure in the lower stratosphere; here ozone absorbs infra-red and solar radiation, causing heating which is mostly balanced by long-wave cooling. On the other hand, dynamical processes such as convection and baroclinic instability (see Section 5.7) are important for heat transport in the troposphere. In the presence of moisture, convection tends to relax the tropospheric lapse rate towards the saturated adiabatic lapse rate (see Section 2.8); the effects of baroclinic instability are more subtle. All of these processes must be taken into account if the position of the tropopause is to be modelled properly; however, although many comprehensive general circulation models produce good simulations of the tropopause height, a complete physical understanding is still lacking.

9.5.2 The middle-atmosphere temperature field

Other aspects of the zonal-mean temperature field also depend on the interplay between dynamics and radiation. It was mentioned in Section 6.6 that there is a large-scale meridional mass circulation in the middle atmosphere, sketched in Figure 6.4, that is driven by wave motions. The full details are beyond the scope of this book; suffice it to say that non-linear and dissipative processes associated with upward-propagating gravity waves drive the summer-to-winter solstitial circulation in the upper mesosphere, including rising motion in the mesosphere over the summer pole and descent over the winter pole, and that non-linear and dissipative processes associated with upward-propagating midlatitude Rossby waves drive the Brewer–Dobson circulation in the stratosphere. As well as being important for tracer transport, as noted in Section 6.6, these circulations also influence the temperature field. It was mentioned in Section 3.6.4 that radiative processes act rather like a ‘spring’, which tries to pull the temperature field towards a purely radiative equilibrium. Such an equilibrium would include warm temperatures in the summer upper mesosphere and near the tropical tropopause, and cool temperatures in the stratospheric polar night. The dynamically driven meridional circulation acts against the radiative spring, forcing temperatures below equilibrium values in the regions of rising motion in the summer upper mesosphere and near the tropical tropopause, and forcing temperatures above equilibrium values in the region of descent in the winter (especially the northern winter) stratosphere.

9.5.3 The Antarctic ozone hole

The Antarctic ozone hole, described in Section 6.7, provides an example of an atmospheric phenomenon in which chemical, dynamical (particularly transport) and radiative effects are

all significant. Dynamically driven transport is of course important for carrying ozone from the ozone-production regions at low latitudes to the polar regions; it also carries chlorine compounds from industrial regions to the Antarctic stratosphere. On the other hand, the polar winter vortex in the Southern Hemisphere is much less disturbed by Rossby wave activity than is the northern winter vortex, so temperatures are closer to the cold radiative-equilibrium values. At such cold temperatures polar stratospheric clouds can form, upon which the crucial heterogeneous chemical reactions that provide the ozone-destroying chlorine compounds take place. Later in the Antarctic spring, when the polar vortex breaks down (owing to dynamical influences), transport may carry ozone-depleted air to lower latitudes, thus perhaps leading to a decrease in the amount of ozone over a larger region than the Antarctic alone.

Further reading

A brief introduction to numerical modelling of the atmosphere is given by Holton (2004) while the book by Kalnay (2002) provides a comprehensive account and also includes a discussion of data assimilation. For an example of the use of distributed computing to study climate change, see Stainforth *et al.* (2005) for some results from the climateprediction.net experiment. Read (1993) gives a critical account of applications of the ideas of chaos theory to the atmosphere and to laboratory analogues of the atmosphere.

Recent work on the processes controlling the height of the midlatitude tropopause is reported by Thuburn and Craig (1997) and Thuburn and Craig (2000). The processes controlling the temperature of the cold equatorial tropopause are discussed by Holton *et al.* (1995). For the summer mesopause see Andrews *et al.* (1987).

A

Appendix A Useful physical constants

In this appendix, STP denotes the standard temperature $273.15\text{ K} = 0^\circ\text{C}$ and pressure $1013.25\text{ hPa} = 1\text{ atm}$. The hectopascal ($\text{hPa} = 10^2\text{ Pa}$) is used as the unit of pressure; it is equivalent to the millibar used in older works. Note that, in SI units, the gramme mole (mol) should strictly speaking be replaced by the kilogramme mole, or kilomole ($\text{kmol} = 10^3\text{ mol}$).

Table A.1

Constant	Symbol	Numerical value
<i>Universal constants</i>		
Universal gas constant or (SI value)	R	$8.31\text{ J K}^{-1}\text{ mol}^{-1}$ $8.31 \times 10^3\text{ J K}^{-1}\text{ kmol}^{-1}$
Avogadro's number or (SI value)	N_A	$6.02 \times 10^{23}\text{ mol}^{-1}$ $6.02 \times 10^{26}\text{ kmol}^{-1}$
Planck constant	h	$6.63 \times 10^{-34}\text{ J s}$
Boltzmann constant	k_B	$1.38 \times 10^{-23}\text{ J K}^{-1}$
Speed of light	c	$3.00 \times 10^8\text{ m s}^{-1}$
Stefan–Boltzmann constant	σ	$5.67 \times 10^{-8}\text{ W m}^{-2}\text{ K}^{-4}$
(Note that $\sigma = 2\pi^5 k^4 / (15h^3 c^2)$)		
<i>The Earth</i>		
Mean acceleration due to gravity at the Earth's surface	g	9.81 m s^{-2}
The Earth's mean radius	a	6371 km
The Earth's mean rate of rotation	Ω	$7.29 \times 10^{-5}\text{ s}^{-1}$
Standard surface pressure	p_0	1013.25 hPa
<i>The Sun</i>		
Solar constant	F_s	1370 W m^{-2}
Mean distance between the Earth and the Sun		$1.50 \times 10^{11}\text{ m}$
Mean radius of the Sun		$6.96 \times 10^8\text{ m}$

(continued)

Table A.1 (cont.)

Constant	Symbol	Numerical value
<i>Dry air</i>		
Molar mass of dry air	M	$28.97 \text{ kg kmol}^{-1}$
Density of dry air at STP	ρ_0	1.29 kg m^{-3}
Specific heat capacity of dry air at STP:		
at constant pressure	c_p	$1005 \text{ J K}^{-1} \text{ kg}^{-1}$
at constant volume	c_v	$718 \text{ J K}^{-1} \text{ kg}^{-1}$
Specific gas constant for dry air	R_d	$287 \text{ J K}^{-1} \text{ kg}^{-1}$
<i>Water</i>		
Molar mass of water	M_w	$18.02 \text{ kg kmol}^{-1}$
Density of liquid water at STP	ρ_w	1000 kg m^{-3}
Density of ice at STP	ρ_i	917 kg m^{-3}
Specific heat capacity of water		
vapour at 0°C :		
at constant pressure		$1850 \text{ J K}^{-1} \text{ kg}^{-1}$
at constant volume		$1390 \text{ J K}^{-1} \text{ kg}^{-1}$
Specific heat capacity of liquid		
water at 0°C		$4217 \text{ J K}^{-1} \text{ kg}^{-1}$
Specific heat capacity of ice at 0°C		$2106 \text{ J K}^{-1} \text{ kg}^{-1}$
Specific gas constant for water vapour	R_v	$461 \text{ J K}^{-1} \text{ kg}^{-1}$
Specific latent heat of vaporization		
at 0°C	L_v	$2.50 \times 10^6 \text{ J kg}^{-1}$
Specific latent heat of vaporization		
at 100°C	L	$2.26 \times 10^6 \text{ J kg}^{-1}$
Specific latent heat of fusion at 0°C	L_f	$0.33 \times 10^6 \text{ J kg}^{-1}$
Specific latent heat of sublimation		
at 0°C	L_s	$2.83 \times 10^6 \text{ J kg}^{-1}$
(Note that $L_s = L_v + L_f$)		
Sources include Kaye and Laby (1986) and Lide (1995) .		

Appendix B Derivation of the equations of motion in spherical coordinates

In this appendix we derive the atmospheric equations of motion (4.21) in spherical coordinates, starting with the Navier–Stokes equation (4.20). In terms of the unit vectors \mathbf{i} , \mathbf{j} and \mathbf{k} , the velocity vector can be written as $\mathbf{u} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ and the rotation vector $\boldsymbol{\Omega} = \Omega(\mathbf{j} \cos \phi + \mathbf{k} \sin \phi)$. Note that the unit vector $\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}/\Omega = \mathbf{j} \cos \phi + \mathbf{k} \sin \phi$ is fixed in space, if the small variations in the Earth’s rotation are neglected.

Now \mathbf{i} , \mathbf{j} and \mathbf{k} change with time, following the motion, and this must be taken into account when calculating the components of $D\mathbf{u}/Dt$; we get

$$\frac{D\mathbf{u}}{Dt} = \frac{Du}{Dt}\mathbf{i} + \frac{Dv}{Dt}\mathbf{j} + \frac{Dw}{Dt}\mathbf{k} + u\frac{Di}{Dt} + v\frac{Dj}{Dt} + w\frac{Dk}{Dt}.$$

The material derivatives of the unit vectors can be calculated as follows. Note first that $\mathbf{k} = \mathbf{r}/r$, where \mathbf{r} is the position vector and r is its magnitude. Then

$$\frac{D\mathbf{k}}{Dt} = \frac{D}{Dt}\left(\frac{\mathbf{r}}{r}\right) = \frac{1}{r}\frac{D\mathbf{r}}{Dt} - \frac{\mathbf{r}}{r^2}\frac{Dr}{Dt}.$$

However, $Dr/Dt = \mathbf{u}$ and $Dr/Dt = w$, so

$$\frac{D\mathbf{k}}{Dt} = \frac{\mathbf{u}}{r} - \frac{\mathbf{k}w}{r} = \frac{1}{r}(u\mathbf{i} + v\mathbf{j}). \quad (\text{B.1})$$

Next use the fact that the unit vector $\hat{\boldsymbol{\Omega}} = \mathbf{j} \cos \phi + \mathbf{k} \sin \phi$ is constant. This implies that

$$0 = \frac{D\hat{\boldsymbol{\Omega}}}{Dt} = \cos \phi \frac{D\mathbf{j}}{Dt} - \frac{v \sin \phi}{r}\mathbf{j} + \sin \phi \frac{D\mathbf{k}}{Dt} + \frac{v \cos \phi}{r}\mathbf{k}, \quad (\text{B.2})$$

since $D\phi/Dt = v/r$. Substitution of equation (B.1) into equation (B.2) then leads to

$$\frac{D\mathbf{j}}{Dt} = -\frac{u \tan \phi}{r}\mathbf{i} - \frac{v}{r}\mathbf{k}. \quad (\text{B.3})$$

Now

$$\mathbf{i} = \mathbf{j} \times \mathbf{k}, \quad (\text{B.4})$$

by orthogonality of the unit vectors. So

$$\frac{Di}{Dt} = \frac{D\mathbf{j}}{Dt} \times \mathbf{k} + \mathbf{j} \times \frac{D\mathbf{k}}{Dt}. \quad (\text{B.5})$$

Substitution from [equations \(B.1\)](#) and [\(B.3\)](#) into [equation \(B.5\)](#) and use of [equation \(B.4\)](#) and the analogous identity $\mathbf{j} = \mathbf{k} \times \mathbf{i}$, together with the identity $\mathbf{k} \times \mathbf{k} = 0$, gives

$$\frac{D\mathbf{i}}{Dt} = \frac{u \tan \phi}{r} \mathbf{j} - \frac{u}{r} \mathbf{k}. \quad (\text{B.6})$$

Alternative, geometrical, derivations of [equations \(B.1\)](#), [\(B.3\)](#) and [\(B.6\)](#) are given in [Section 2.3 of Holton \(2004\)](#).

The Coriolis term in [equation \(4.20\)](#) can be written

$$2\Omega \times \mathbf{u} = 2\Omega(w \cos \phi - v \sin \phi) \mathbf{i} + 2\Omega u \sin \phi \mathbf{j} - 2\Omega u \cos \phi \mathbf{k}.$$

In terms of the small incremental distances $dx = r \cos \phi d\lambda$ in the eastward direction, $dy = rd\phi$ in the northward direction and $dz = dr$ in the vertical direction, the pressure gradient term in [equation \(4.20\)](#) can be written

$$\nabla p = \frac{\partial p}{\partial x} \mathbf{i} + \frac{\partial p}{\partial y} \mathbf{j} + \frac{\partial p}{\partial z} \mathbf{k}.$$

Neglecting the centripetal acceleration, [equation \(4.20\)](#) can therefore be written

$$\begin{aligned} & \frac{Du}{Dt} \mathbf{i} + \frac{Dv}{Dt} \mathbf{j} + \frac{Dw}{Dt} \mathbf{k} \\ & + \frac{u}{r}(u \tan \phi \mathbf{j} - u \mathbf{k}) - \frac{v}{r}(u \tan \phi \mathbf{i} + v \mathbf{k}) + \frac{w}{r}(u \mathbf{i} + v \mathbf{j}) \\ & + 2\Omega(w \cos \phi - v \sin \phi) \mathbf{i} + 2\Omega u \sin \phi \mathbf{j} - 2\Omega u \cos \phi \mathbf{k} \\ & + \frac{1}{\rho} \left(\frac{\partial p}{\partial x} \mathbf{i} + \frac{\partial p}{\partial y} \mathbf{j} + \frac{\partial p}{\partial z} \mathbf{k} \right) + g \mathbf{k} = \mathbf{F}, \end{aligned}$$

where $\mathbf{F} = F^{(x)} \mathbf{i} + F^{(y)} \mathbf{j} + F^{(z)} \mathbf{k}$ is the frictional force. When the frictional force is provided by molecular viscosity alone, then $\mathbf{F} = \mathbf{F}_{\text{visc}}$. Collecting the terms in \mathbf{i} , \mathbf{j} and \mathbf{k} , we then get [equation \(4.21a\)](#), [equation \(4.21b\)](#) and [equation \(4.21c\)](#), respectively, as required.

References

- Acheson, D. J. (1990). *Elementary Fluid Dynamics*. Oxford University Press.
- Andrews, D. G. (1991). A stratospheric transport system. *Phys. World*, **4**(11), 41–46.
- Andrews, D. G., Holton, J. R., and Leovy, C. B. (1987). *Middle Atmosphere Dynamics*. Academic Press.
- Archer, D. (2007). *Global Warming: Understanding the Forecast*. Blackwell Publishing.
- Atkins, P. W. (2006). *Physical Chemistry*. Oxford University Press, 8th edition.
- Balsley, B. B. and Gage, K. S. (1980). The MST radar technique: potential for middle atmospheric studies. *Pure Appl. Geophys.*, **118**, 452–493.
- Balsley, B. B., Ecklund, W. L., and Fritts, D. C. (1983). VHF echoes from the high-latitude mesosphere and lower thermosphere: observations and interpretations. *J. Atmos. Sci.*, **40**, 2451–2466.
- Banwell, C. N. and McCash, E. M. (1994). *Fundamentals of Molecular Spectroscopy*. McGraw-Hill, 4th edition.
- Bates, R. (2007). Some considerations of the concept of climate feedback. *Quart. J. Roy. Meteorol. Soc.*, **133**, 545–560.
- Blundell, S. J. and Blundell, K. M. (2009). *Concepts in Thermal Physics*. Oxford University Press, 2nd edition.
- Boas, M. L. (1983). *Mathematical Methods in the Physical Sciences*. Wiley, 2nd edition.
- Bohren, C. F. (1987). *Clouds in a Glass of Beer*. Wiley.
- Bohren, C. F. and Albrecht, B. A. (1998). *Atmospheric Thermodynamics*. Oxford University Press.
- Bolton, D. (1980). The computation of equivalent potential temperature. *Mon. Weather Rev.*, **108**, 1046–1053.
- Born, M. and Wolf, E. (1980). *Principles of Optics*. Pergamon, 6th edition.
- Bransden, B. H. and Joachain, C. J. (1989). *Introduction to Quantum Mechanics*. Longman.
- Brasseur, G. and Solomon, S. (2005). *Aeronomy of the Middle Atmosphere*. Springer, 3rd edition.
- Brewer, A. W. (1949). Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quart. J. Roy. Meteorol. Soc.*, **75**, 351–363.
- Chapman, S. (1930). A theory of upper atmospheric ozone. *Mem. Roy. Meteorol. Soc.*, **3**, 103–125.
- Charney, J. G. and Drazin, P. G. (1961). Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83–109.
- DeMore, W. B., Sander, S. P., Golden, D. M., Hampson, R. F., Kurylo, M. J., Howard, C. J., Ravishankara, A. R., Kolb, C. E., and Molina, M. J. (1997). Chemical kinetics

- and photochemical data for use in stratospheric modeling, Evaluation Number 12. JPL Publication 97-4, Jet Propulsion Laboratory.
- Dessler, A. E. (2000). *The Chemistry and Physics of Stratospheric Ozone*. Academic Press.
- Dobson, G. M. B. (1956). Origin and distribution of the polyatomic molecules in the atmosphere. *Proc. Roy. Soc., A* **236**, 187–193.
- Dobson, G. M. B. and Normand, C. W. B. (1957). Observers' handbook for the ozone spectrophotometer. *Ann. Int. Geophys. Year*, **5**, 46–81.
- Durran, D. R. (1993). Is the Coriolis force really responsible for the inertial oscillation? *Bull. Am. Meteorol. Soc.*, **74**, 2179–2184.
- Faber, T. E. (1995). *Fluid Dynamics for Physicists*. Cambridge University Press.
- Fleming, E. L., Chandra, S., Barnett, J. J., and Corney, M. (1990). Zonal mean temperature, pressure, zonal wind and geopotential height as functions of latitude. *Adv. Space Res.*, **10**(12), 11–59. For data tables of the COSPAR Middle Atmosphere Reference Atmosphere, see pages 357–517.
- Gardiner, B. G. (1989). The Antarctic ozone hole. *Weather*, **44**, 291–298.
- Gill, A. E. (1982). *Atmosphere–Ocean Dynamics*. Academic Press.
- Glazer, A. M. and Wark, J. S. (2001). *Statistical Mechanics: A Survival Guide*. Oxford University Press.
- Goody, R. M. (1995). *Principles of Atmospheric Physics and Chemistry*. Oxford University Press.
- Goody, R. M. and Yung, Y. L. (1989). *Atmospheric Radiation*. Oxford University Press, 2nd edition.
- Graedel, T. E. and Crutzen, P. J. (1995). *Atmosphere, Climate, and Change*. Scientific American Library.
- Grant, I. S. and Phillips, W. R. (1990). *Electromagnetism*. Wiley, 2nd edition.
- Hanel, R. A., Schlachman, B., Rogers, D., and Vanous, D. (1971). Nimbus 4 Michelson interferometer. *Appl. Opt.*, **10**, 1376–1382.
- Hanel, R. A., Conrath, B., Hovis, W., Kunde, V., Lowman, P., Maguire, W., Pearl, J., Pirraglia, J., Prabhakara, C., Schlachman, B., Levin, G., Straat, P., and Burke, T. (1972). Investigation of the Martian environment by infrared spectroscopy on Mariner 9. *Icarus*, **17**, 423–442.
- Hanel, R. A., Conrath, B. J., Jennings, D. E., and Samuelson, R. E. (2003). *Exploration of the Solar System by Infrared Remote Sensing*. Cambridge University Press, 2nd edition.
- Hartmann, D. L. (1994). *Global Physical Climatology*. Academic Press.
- Herzberg, G. (1945). *Molecular Spectra and Molecular Structure. II. Infra-red and Raman Spectra of Polyatomic Molecules*. Van Nostrand.
- Herzberg, L. (1965). Solar optical radiation and its role in upper atmospheric processes. In C. O. Hines, I. Paglis, T. R. Hartz, and J. A. Fejer, editors, *Physics of the Earth's Upper Atmosphere*, pages 31–45. Prentice Hall.
- Hide, R. and Mason, P. J. (1975). Sloping convection in a rotating fluid. *Adv. Phys.*, **24**, 47–100.
- Hobbs, P. V. (1995). *Basic Physical Chemistry for the Atmospheric Sciences*. Cambridge University Press.

- Holton, J. R. (2004). *An Introduction to Dynamic Meteorology*. Academic Press, 4th edition.
- Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L. (1995). Stratosphere–troposphere exchange. *Rev. Geophys.*, **33**, 403–439.
- Houghton, J. T. (2002). *The Physics of Atmospheres*. Cambridge University Press, 3rd edition.
- Houghton, J. T. (2009). *Global Warming: The Complete Briefing*. Cambridge University Press, 4th edition.
- Houghton, J. T., Taylor, F. W., and Rodgers, C. D. (1984). *Remote Sounding of Atmospheres*. Cambridge University Press.
- Jackson, J. D. (1998). *Classical Electrodynamics*. Wiley, 3rd edition.
- Kalnay, E. (2002). *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.
- Kaye, G. W. C. and Laby, T. H. (1986). *Tables of Physical and Chemical Constants*. Longman, 15th edition.
- Landau, L. D. and Lifshitz, E. M. (1980). *Statistical Physics, Part 1*. Pergamon, 3rd edition.
- Lary, D. J. (1997). Catalytic destruction of stratospheric ozone. *J. Geophys. Res.*, **102**, 21515–21526.
- Li, D. and Shine, K. P. (1995). A 4-dimensional ozone climatology for UGAMP models. UK Universities' Global Atmospheric Modelling Programme Internal Report No. 35.
- Lide, D. R., editor (1995). *Handbook of Chemistry and Physics*. CRC Press, 76th edition.
- Liou, K. (2002). *An Introduction to Atmospheric Radiation*. Academic Press, 2nd edition.
- London, J. (1980). Radiative energy sources and sinks in the stratosphere and mesosphere. In A. Aiken, editor, *Proceedings of the NATO Advanced Study Institute on Atmospheric Ozone: Its Variation and Human Influences*, pages 703–721. Report FAA-EE-80-20.
- Lorenz, E. N. (1955). Available potential energy and the maintenance of the general circulation. *Tellus*, **7**, 157–167.
- Lorenz, E. N. (1993). *The Essence of Chaos*. UCL Press.
- Lorrain, P., Corson, D. P., and Lorrain, F. (1988). *Electromagnetic Fields and Waves*. Freeman, 3rd edition.
- Lynch, D. K. and Livingston, W. (2001). *Color and Light in Nature*. Cambridge University Press, 2nd edition.
- Lyons, L. (1995). *All You Wanted to Know About Mathematics But Were Afraid to Ask*, volume 1. Cambridge University Press.
- Mandl, F. (1988). *Statistical Physics*. Wiley, 2nd edition.
- Marshall, J. and Plumb, R. (2008). *Atmosphere, Ocean and Climate Dynamics: An Introductory Text*. Academic Press.
- McIlveen, R. (1991). *Fundamentals of Weather and Climate*. Chapman and Hall.
- McIntosh, D. H. and Thom, A. S. (1983). *Essentials of Meteorology*. Taylor and Francis.
- Meier, R. R. (1991). Ultraviolet spectroscopy and remote sensing of the upper atmosphere. *Space Sci. Rev.*, **58**, 1–185.
- Myhre, G., Highwood, E. J., Shine, K. P., and Stordal, F. (1998). New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.*, **25**, 2715–2718.
- Pedlosky, J. (1987). *Geophysical Fluid Dynamics*. Springer-Verlag, 2nd edition.

- Petty, G. W. (2004). *A First Course in Atmospheric Radiation*. Sundog Publishing.
- Pierrehumbert, R. T. (2010). *Principles of Planetary Climate*. Cambridge University Press. (In press).
- Pierrehumbert, R. T., Brogniez, H., and Roca, R. (2007). On the relative humidity of the atmosphere. In T. Schneider and A. Sobel, editors, *The Global Circulation of the Atmosphere*, pages 143–185. Princeton University Press.
- Rae, A. I. M. (2007). *Quantum Mechanics*. Taylor and Francis, 5th edition.
- Ramanathan, V. and Dickinson, R. E. (1979). The role of stratospheric ozone in the zonal and seasonal radiative energy balance of the Earth-troposphere system. *J. Atmos. Sci.*, **36**, 1084–1104.
- Read, P. L. (1993). Applications of chaos to meteorology and climate. In T. Mullin, editor, *The Nature of Chaos*, pages 222–260. Oxford University Press.
- Riley, K. F., Hobson, M. P., and Bence, S. J. (2006). *Mathematical Methods for Physics and Engineering: A Comprehensive Guide*. Cambridge University Press.
- Rodgers, C. D. and Walshaw, C. D. (1966). The computation of infra-red cooling rate in planetary atmospheres. *Quart. J. Roy. Meteorol. Soc.*, **92**, 67–92.
- Rogers, R. R. and Yau, M. K. (1989). *A Short Course in Cloud Physics*. Pergamon, 3rd edition.
- Salby, M. L. (1996). *Fundamentals of Atmospheric Physics*. Academic Press.
- Solomon, S. (1999). Stratospheric ozone depletion: A review of concepts and history. *Rev. Geophys.*, **37**, 275–316.
- Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., editors (2007). *IPCC, 2007: Climate Change 2007: The Physical Science Basis*. Cambridge University Press. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. See also <http://www.ipcc.ch/>.
- Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.
- Thomas, L. (1987). Laser radar observations of middle-atmosphere structure and composition. *Phil. Trans. Roy. Soc.*, **A323**, 597–609.
- Thorne, A. P. (1988). *Spectrophysics*. Chapman and Hall, 2nd edition.
- Thuburn, J. and Craig, G. C. (1997). GCM tests of theories for the height of the tropopause. *J. Atmos. Sci.*, **54**, 869–882.
- Thuburn, J. and Craig, G. C. (2000). Stratospheric influence on tropopause height: the radiative constraint. *J. Atmos. Sci.*, **57**, 17–28.
- Tritton, D. J. (1988). *Physical Fluid Dynamics*. Oxford University Press, 2nd edition.
- US Standard Atmosphere (1976). National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, US Air Force. US Government Printing Office.
- Vallis, G. K. (2006). *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press.

- Vaughan, G., Wareing, D.P., Jones, S.B., Thomas, L., and Larsen, N. (1994). Lidar measurements of Mt. Pinatubo aerosols at Aberystwyth from August 1991 through March 1992. *Geophys. Res. Lett.*, **21**, 1315–1318.
- Wallace, J.M. and Hobbs, P.V. (2006). *Atmospheric Science: An Introductory Survey*. Academic Press, 2nd edition.
- Wayne, R.P. (2000). *Chemistry of Atmospheres*. Oxford University Press, 3rd edition.

Index

- absorptance, 6, 70, 72, 81, 173, 175
spectral, 56, 62, 175
- absorption
band, 73–74
coefficient, 60, 73, 207
continuum, 73
cross-section, 73, 74, 157
line, 175
- activated droplet, 46
- activation energy, 155
- adiabatic process, 26, 27
- adiathermal process, 26
- aerosol, 53, 188, 189, 196, 202
- albedo
planetary, 5, 198, 213
single scattering, 87, 93
- backscatter, 86, 180–181, 183, 188, 189
- Beer’s law, 61
- β -plane, 107
- black body, 5, 6, 55, 81, 85, 173, 205, 207
- Boltzmann distribution, 56
- boundary layer, 136–149, 219
- Boussinesq approximation, 122, 123, 125, 131, 137, 141
- Boussinesq equations, 124, 143
linearised, 125, 128
- Brewer–Dobson circulation, 163, 164, 223
- brightness temperature, 174, 178
- Brunt–Väisälä frequency, *see* buoyancy frequency
- buoyancy force, 2, 29, 132
- buoyancy frequency, 30, 123, 131
- carbon dioxide, 1, 10, 17, 21, 64, 72, 80, 168, 179, 196, 208
- catalytic cycle, 160, 166, 169
- centrifugal force, 104, 110
- centripetal acceleration, 103
- chaos, 17
- Chapman layer, 75, 77, 80, 158, 178
- Chapman ozone chemistry, 158, 160, 169
- Charney–Drazin criterion, 134
- chemical kinetics, 153
- chemical lifetime, 153, 162
- chlorofluorocarbons (CFCs), 16, 167, 211
- Clausius–Clapeyron equation, 33, 36, 38, 197, 206
- climate, 17
change, 3
feedback parameter, 199
forecasting, 4, 219
sensitivity, 200, 213
system, 198
heat capacity of, 198
- cloud condensation nucleus, 35, 42, 45
- clouds, 1, 42, 45, 195, 218, 219
polar stratospheric (PSCs), 32, 165, 167, 224
- column ozone, 15, 164, 182
- continuity equation, *see* mass conservation law
- convection, 10, 86, 223
- cooling to space, 78, 92
- cooling, long-wave, 10, 80
- Coriolis acceleration, 103
- Coriolis force, 2, 3, 104, 108, 110, 115
- Coriolis parameter, 105
- Curtis–Godson approximation, 92
- data assimilation, 4, 220
- dew point, 36, 42
- diabatic heating, 75, 113, 152
- diffuse approximation, 62, 83
- dispersion relation
Eady waves, 145
inertia–gravity waves, 148
internal gravity waves, 130
Rossby waves, 133, 148
- Dobson ozone spectrophotometer, 164, 182, 183
- Dobson Unit (DU), 164
- dry adiabatic curve, 40
- dynamical heating, 80
- effective emitting temperature, 5, 81, 85
- effective gravity, 104, 115
- Ekman
layer, 140, 149
pumping velocity, 139, 140
spiral, 140, 141, 149
volume transport, 137
- electromagnetic waves, 52
- emission
height, 85
line, 175

- of photon, 54
 thermal, 173
 emittance, 6
 spectral, 55, 56, 62, 175
 stratospheric, 81
 energy balance model (EBM), 198, 199, 212–213
 energy equation, 125
 enthalpy, 25, 31, 48
 of reaction, 152
 of vaporisation, 33
 entropy, 2, 25, 26, 33, 40, 113
 equation of state, 99
 equivalent width, 70
 Eulerian picture of fluid flow, 96, 114
 extinction, 53
 coefficient, 60, 66, 67, 76, 83, 87, 178, 179
- f*-plane, 107
 feedback, 204, 206
 black-body, 205, 213
 ice-albedo, 206, 213
 negative, 206
 positive, 197, 206
 response time (FRT), 200
 water vapour, 206
 frictional stress, 136
 front, atmospheric, 116
 frost point, 37
- gas constant, 20
 general circulation model (GCM), 4, 215, 223
 geopotential, 112
 geostrophic
 approximation, 108, 109
 balance, 108, 109, 112, 125
 flow, 126–128, 135, 141
 Gibbs free energy, 42–44, 152, 153
 gradient–wind balance, 111
 gravity wave, 3, 12, 128, 131, 164, 219, 223
 inertia–, 13, 132, 148, 188
 internal, 131, 148
 greenhouse effect, 4, 5, 7, 81, 83, 195, 212
 enhanced, 17, 196
 greenhouse gas, 7, 65, 195
 grey absorber, 83, 212
 group velocity, 13–14, 130, 134, 135
- haze droplet, 46
 heating rate, radiative, 75, 77–80, 92, 113
 hierarchy of models, 4, 215
 hydrostatic balance, 1, 9, 22, 108, 111, 130
- ice particles, 16, 163
 ice phase, 36
 ideal gas, 2, 19, 20, 23, 99, 113
- incompressible fluid, 99, 123
 inertial oscillation, 115, 147
 instability
 baroclinic, 144, 223
 Eady's model, 144, 146, 150
 barotropic, 147
 conditional, 39
 convective, 142
 numerical, 218
 static, 10, 29, 30, 141
 Intergovernmental Panel on Climate Change (IPCC), 17, 195
 internal energy, 25, 31, 113
 irradiance, 4, 58
 black-body, 63, 71, 195, 207
 spectral, 58, 59, 62, 63, 71, 76, 78, 158, 207
 isentrope, 32
 isobar, 2, 41, 109, 117, 141
 isotherm, 40, 109
- Köhler curve, 46
 Kelvin's formula, 44
 kinetic energy, 31, 32, 125, 144, 148
 molecular, 2, 53, 54, 57, 157
 Kirchhoff's law, 6, 56, 61
- laboratory models of the atmosphere, 220
 Lagrangian picture of fluid flow, 97–99, 113, 115
 Lambert's law, 60
 lapse rate, 24, 27, 48, 49, 86, 142, 211, 223
 dry adiabatic (DALR), 27, 28
 saturated adiabatic (SALR), 10, 38
 latent heat, 2, 11, 19, 33, 36–38, 50, 113
 lidar, 188, 189
 lifting condensation level, 35, 50
 limb viewing, 172, 179, 190
 linearisation, 124, 133, 137, 144, 146, 199
 local thermodynamic equilibrium (LTE), 57, 61, 78
- lower atmosphere, *see troposphere*
- mass conservation law, 95, 98, 112, 162
 material derivative, 97, 107
 mesopause, 8, 10, 11, 152
 mesosphere, 8, 12, 13, 80
 circulation of, 164, 223
 middle atmosphere, 8
 diabatic heating of, 80, 113
 meridional circulation, 223
 mixing ratio, 21, 161–163, 206
 mass, 20, 21, 34, 161
 saturation, 35, 37, 41, 206
 volume, 15, 21
 moisture, 32–47, 163, 219
 molar mass, 21

- nadir viewing, 172, 173, 179, 190
 Navier–Stokes equation, 99, 101, 102, 104
 Newtonian cooling, 80
 nitrogen, 1, 21, 64, 72, 152
 number density, 15, 21, 153, 161–163,
 188, 189
 ozone, 158, 160
 numerical methods, 217
 finite difference, 217
 spectral, 218
 numerical weather prediction (NWP),
 219, 220
- odd oxygen, 158–160
 optical depth, 62, 75–77, 83–85, 178, 180, 181
 optical path, 61, 76, 182, 188
 outgoing long-wave radiation (OLR), 196,
 207–211
 oxygen, 1, 8, 21, 72–74, 80, 152, 158, 159
 ozone, 1, 8–10, 14, 15, 21, 64, 72–74, 80,
 158–161, 163, 164, 166, 168, 182, 192, 196,
 216, 223
 depletion of, 16, 167
 hole, Antarctic, 16, 164, 165, 181, 220, 223
 layer, 8, 77, 158, 164
 transport of, 163
- parameterisation, 219
 parcel of air, 15, 26–30, 141, 151, 152, 155,
 161–163
 displacement in wave, 130, 131, 148
 interchange, 143, 150
 moist, 34, 35, 37–39
 partial pressure, 20, 21, 34, 42–44, 186
 partial reflection radar, 186
 phase changes, 2, 33
 phase speed, 14
 photo-dissociation, 53, 73, 74, 77, 157, 167
 photo-ionisation, 53, 73
 photolysis, *see* photo-dissociation
 photon, 52
 absorption of, 8, 53
 long-wave, *see* photon, thermal
 short-wave, *see* photon, solar
 solar, 1, 53, 65, 152, 157
 thermal, 53, 54, 63, 78
 Planck function, 54, 56–58, 173–176, 178, 207,
 209, 210
 Planck’s law, 54, 57
 planetary wave, *see* Rossby wave
 potential energy, 30, 65, 143, 144, 155
 available, 32, 49, 125
 total, 31, 32
 potential temperature, 26, 27, 30, 35, 114
 equivalent, 40, 41
 wet-bulb, 41
- potential vorticity, 121, 122, 128, 161
 quasi-geostrophic (QGPV), 128, 133, 135, 145
 inversion, 128, 135
 Poynting vector, 59, 125
 pressure coordinates, 111, 116, 177
- quantum yield, 157
 quasi-biennial oscillation, 12
 quasi-geostrophic
 equations, 126, 133
 flow, 144
- radar, 12, 128, 183
 Doppler shift, 187
- radiance, 58
 spectral, 54, 55, 57, 173, 180, 182
- radiation
 black-body, 54
 infra-red
 absorption of, 72
 far, 52
 near, 52
 thermal, 52
 long-wave, *see* radiation, thermal
 short-wave, *see* radiation, solar
 solar, 3, 4, 6, 7, 53, 80, 172
 absorption of, 7, 10, 76
 thermal, 6, 53, 172
 absorption of, 7
 ultra-violet, 9
 absorption of, 73
 radiative equilibrium, 6, 80, 81, 83, 84
 radiative forcing, 196, 199–202, 212
 due to carbon dioxide, 202, 207, 211, 213
 radiative-transfer equation, 61, 69, 78, 83,
 87, 173
 range resolution, 185, 186, 189
 Raoult’s law, 45
 rate coefficient, 153, 154, 156, 159
 Rayleigh friction, 149
 reaction
 chemical, 151, 153
 bimolecular, 154, 155, 168
 heterogeneous, 166, 224
 photochemical, 1, 8, 15, 53, 157, 216
 relative humidity, 34, 42–46
 remote sounding, 3, 171, 173, 179, 182
 Rossby number, 109, 115, 126, 133
 Rossby wave, 3, 13, 132, 134–136, 221,
 223, 224
 stationary, 13, 134
 rotating annulus, 221
 rotating reference frame, 102–104
 rotational
 state, 54, 63, 64
 transition, 64, 66, 72, 73

- satellite observations, 164, 172–181
 saturated adiabatic curve, 39, 41
 saturation vapour pressure (SVP), 33, 34
 scale analysis, 107
 scale height
 density, 122
 pressure, 23, 49
 water vapour, 49
 scattering
 cross-section, differential, 185
 Mie, 53
 Rayleigh, 53, 180, 189
 simple model of, 86
 Schwarzschild's equation, *see* radiative-transfer equation
 skin temperature, 85
 sloping convection, *see* instability, baroclinic
 solar constant, *see* total solar irradiance
 solar spectrum, 71
 source function, 60
 specific humidity, 20, 34
 spectral band, 64, 70, 72, 208
 spectral line, 64, 66
 broadening, 67, 68
 Doppler, 68, 91
 Lorentz, 67, 90, 91
 strength, 66
 strong, 70, 91
 weak, 70, 91
 spherical coordinates, 104
 static stability, 28, 30
 Stefan–Boltzmann law, 5, 59, 205
 stratification, density, 2, 49, 123, 144
 stratosphere, 8, 9
 stratosphere, 8, 12, 81
 ozone in, 3, 158–161
 temperature of, 83
 stratospheric warming, 13
 stretching term, 127
 sublimation, 36, 37, 51
 synoptic-scale systems, 107
 tangent plane, 106
 tephigram, 40, 50
 thermal windshear balance, 11, 109, 112, 116
 thermodynamic energy equation, *see* thermodynamics, first law of
 thermodynamic equilibrium, 54
 Thermodynamics, First Law of, 24, 113, 151
 thermosphere, 8
 thickness between pressure surfaces, 24
 total ozone, *see* column ozone
 Total Ozone Mapping Spectrometer (TOMS), 164, 181
 total solar irradiance (TSI), 5, 198
 tracer of atmospheric motion, 122, 161
 transmittance, 6, 69, 173, 207, 212
 transport of chemicals, 21, 161, 216
 triple point, 36
 tropopause, 8, 9, 11, 163, 223
 troposphere, 8, 10
 tropical, 10, 147
 two-stream approximation, 84
 upper atmosphere, 9
 vibrational
 state, 54, 63, 64, 66
 transition, 64, 66, 72–74
 viscosity, 101, 221
 eddy, 136, 139, 141
 vorticity, 119–121, 135
 absolute, 121, 127
 relative, 121
 vorticity equation, 127
 water vapour, 1, 32–34, 36–38, 42, 43, 46, 47, 64, 72, 80, 113, 187, 196, 197, 204, 205
 diffusion of, 46
 weather, 8, 17
 forecasting, 17, 219
 ensemble, 219
 map, 117
 weighting function, 178–181, 193
 window, atmospheric, 72, 192