

of the PHYSICS
ATMOSPHERE
and CLIMATE

MURRY L. SALBY

Physics of the Atmosphere and Climate

Murry Salby's new book provides an integrated treatment of the processes controlling the Earth-atmosphere system developed from first principles through a balance of theory and applications. This book builds on Salby's previous book *Fundamentals of Atmospheric Physics*. The scope has been expanded to include climate, while streamlining the presentation for undergraduates in science, mathematics, and engineering. Advanced material, suitable for graduate students and researchers, has been retained but distinguished from the basic development. The book offers a conceptual yet quantitative understanding of the controlling influences integrated through theory and major applications. It leads readers through a methodical development of the diverse physical processes that shape weather, global energetics, and climate. End-of-chapter problems of varying difficulty develop student knowledge and its quantitative application, supported by answers and detailed solutions online for instructors.

MURRY SALBY is Chair of Climate Science at Macquarie University, Sydney, Australia. He was previously a Professor at the University of Colorado, where he served as Director of the Center for Atmospheric Theory and Analysis. Previously, he was a researcher at the U.S. National Center for Atmospheric Research and at Princeton University. Professor Salby has authored more than 100 scientific articles in major international journals, as well as the textbook *Fundamentals of Atmospheric Physics* (1996). His research focuses on changes of the atmospheric circulation in relation to global structure, energetics, and climate. Involving large-scale computer simulation and satellite data, Salby's research has provided insight into a wide range of phenomena in the Earth-atmosphere system.

Praise for *Physics of the Atmosphere and Climate*

“Salby’s book is a graduate textbook on Earth’s atmosphere and climate that is well balanced between the physics of the constituent materials and fluid dynamics. I recommend it as a foundation for anyone who wants to do research on the important open questions about aerosols, radiation, biogeochemistry, and ocean-atmosphere coupling.”

-Professor Jim McWilliams, University of California, Los Angeles

“Salby’s book provides an exhaustive survey of the atmospheric and climate sciences. The topics are well motivated with thorough discussion and are supported with excellent figures. The book is an essential reference for researchers and graduate and advanced undergraduate students who wish to have a rigorous source for a wide range of fundamental atmospheric science topics. Each chapter ends with an excellent selection of additional references and a challenging set of problems. Atmospheric and climate scientists will find this book to be an essential one for their libraries.”

-Associate Professor Hampton N. Shirer, Pennsylvania State University

“Murry Salby presents an informative and insightful tour through the contemporary issues in the atmospheric sciences as they relate to climate. *Physics of the Atmosphere and Climate* is a valuable resource for educators and researchers alike, serving both as a textbook for the graduate or advanced undergraduate student with a physics or mathematics background and as an excellent reference and refresher for practitioners. It is a welcome addition to the field.”

-Professor Darin W. Toohey, University of Colorado at Boulder

Salby’s earlier book is a classic. As a textbook it is unequalled in breadth, depth, and lucidity. It is the single volume that I recommend to all of my students in atmospheric science. This new version improves over the previous version, if that is possible, in three aspects: beautiful illustrations of global processes (e.g. hydrological cycle) from newly available satellite data, new topics of current interest (e.g. interannual changes in the stratosphere and the oceans), and a new chapter on the influence of the ocean on the atmosphere. These changes make the book more useful as a starting point for studying climate change.”

-Professor Yuk Yung, California Institute of Technology

PHYSICS OF THE ATMOSPHERE AND CLIMATE

MURRY L. SALBY
Macquarie University



CAMBRIDGE
UNIVERSITY PRESS

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

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA
www.cambridge.org
Information on this title: www.cambridge.org/9780521767187

© Murry L. Salby 2012

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

First published 2012

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data

Salby, Murry L.

Physics of the atmosphere and climate / Murry L. Salby. - 2nd ed.

p. cm.

Revised ed. of: *Fundamentals of atmospheric physics*. 1996.

Includes bibliographical references and index.

ISBN 978-0-521-76718-7 (hardback)

1. Atmospheric physics. I. Salby, Murry L. *Fundamentals of atmospheric physics*. II. Title.

QC861.3.S257 2011

551.5-dc23 2011033950

ISBN 978-0-521-76718-7 Hardback

Additional resources for this publication at www.cambridge.org/salby

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

To Alon

Strive for the Best

Contents

Preface	<i>page</i> xv
Prelude	xix
1 The Earth-atmosphere system	1
1.1 Introduction	1
1.1.1 Descriptions of atmospheric behavior	1
1.1.2 Mechanisms influencing atmospheric behavior	2
1.2 Composition and structure	3
1.2.1 Description of air	3
1.2.2 Stratification of mass	7
1.2.3 Thermal and dynamical structure	14
1.2.4 Trace constituents	20
1.2.5 Cloud	41
1.3 Radiative equilibrium of the Earth	46
1.4 The global energy budget	48
1.4.1 Global-mean energy balance	48
1.4.2 Horizontal distribution of radiative transfer	50
1.5 The general circulation	53
1.6 Historical perspective: Global-mean temperature	56
1.6.1 The instrumental record	57
1.6.2 Proxy records	65
Suggested references	71
Problems	71
2 Thermodynamics of gases	74
2.1 Thermodynamic concepts	74
2.1.1 Thermodynamic properties	75
2.1.2 Expansion work	75
2.1.3 Heat transfer	77
2.1.4 State variables and thermodynamic processes	77

2.2	The First Law	81
2.2.1	Internal energy	81
2.2.2	Diabatic changes of state	82
2.3	Heat capacity	83
2.4	Adiabatic processes	86
2.4.1	Potential temperature	87
2.4.2	Thermodynamic behavior accompanying vertical motion	90
2.5	Diabatic processes	91
2.5.1	Polytropic processes	92
	Suggested references	93
	Problems	93
3	The Second Law and its implications	95
3.1	Natural and reversible processes	95
3.1.1	The Carnot cycle	98
3.2	Entropy and the Second Law	99
3.3	Restricted forms of the Second Law	102
3.4	The fundamental relations	103
3.4.1	The Maxwell Relations	104
3.4.2	Noncompensated heat transfer	104
3.5	Conditions for thermodynamic equilibrium	105
3.6	Relationship of entropy to potential temperature	107
3.6.1	Implications for vertical motion	107
	Suggested references	110
	Problems	110
4	Heterogeneous systems	112
4.1	Description of a heterogeneous system	112
4.2	Chemical equilibrium	115
4.3	Fundamental relations for a multi-component system	117
4.4	Thermodynamic degrees of freedom	118
4.5	Thermodynamic characteristics of water	119
4.6	Equilibrium phase transformations	122
4.6.1	Latent heat	122
4.6.2	Clausius-Clapeyron Equation	123
	Suggested references	125
	Problems	126
5	Transformations of moist air	127
5.1	Description of moist air	127
5.1.1	Properties of the gas phase	127
5.1.2	Saturation properties	129
5.2	Implications for the distribution of water vapor	131
5.3	State variables of the two-component system	132
5.3.1	Unsaturated behavior	133
5.3.2	Saturated behavior	133
5.4	Thermodynamic behavior accompanying vertical motion	135
5.4.1	Condensation and the release of latent heat	135

5.4.2	The pseudo-adiabatic process	139
5.4.3	The Saturated Adiabatic Lapse Rate	141
5.5	The pseudo-adiabatic chart	142
	Suggested references	146
	Problems	146
6	Hydrostatic equilibrium	150
6.1	Effective gravity	150
6.2	Geopotential coordinates	152
6.3	Hydrostatic balance	153
6.3.1	Hypsometric equation	154
6.3.2	Meteorological Analyses	154
6.4	Stratification	157
6.4.1	Idealized stratification	159
6.5	Lagrangian interpretation of stratification	161
6.5.1	Adiabatic stratification: A paradigm of the troposphere	162
6.5.2	Diabatic stratification: A paradigm of the stratosphere	165
	Suggested references	168
	Problems	168
7	Static stability	171
7.1	Reaction to vertical displacement	171
7.2	Stability categories	173
7.2.1	Stability in terms of temperature	174
7.2.2	Stability in terms of potential temperature	175
7.2.3	Moisture dependence	178
7.3	Implications for vertical motion	179
7.4	Finite displacements	180
7.4.1	Conditional instability	181
7.4.2	Entrainment	186
7.4.3	Potential instability	188
7.4.4	Modification of stability under unsaturated conditions	190
7.5	Stabilizing and destabilizing influences	191
7.6	Turbulent dispersion	194
7.6.1	Convective mixing	194
7.6.2	Inversions	194
7.6.3	Life cycle of the nocturnal inversion	195
7.7	Relationship to observed thermal structure	197
	Suggested references	200
	Problems	200
8	Radiative transfer	203
8.1	Shortwave and longwave radiation	203
8.1.1	Spectra of observed SW and LW radiation	204
8.2	Description of radiative transfer	209
8.2.1	Radiometric quantities	209

8.2.2	Absorption	211
8.2.3	Emission	212
8.2.4	Scattering	216
8.2.5	The Equation of Radiative Transfer	218
8.3	Absorption characteristics of gases	219
8.3.1	Interaction between radiation and molecules	219
8.3.2	Line broadening	224
8.4	Radiative transfer in a plane parallel atmosphere	227
8.4.1	Transmission function	229
8.4.2	Two-stream approximation	231
8.5	Thermal equilibrium	234
8.5.1	Radiative equilibrium in a gray atmosphere	234
8.5.2	Radiative-convective equilibrium	237
8.5.3	Radiative heating	240
8.6	Thermal relaxation	245
8.7	The greenhouse effect	247
8.7.1	Feedback in the climate system	249
8.7.2	Unchecked feedback	258
8.7.3	Simulation of climate	260
	Suggested references	262
	Problems	262
9	Aerosol and cloud	266
9.1	Morphology of atmospheric aerosol	266
9.1.1	Continental aerosol	266
9.1.2	Marine aerosol	271
9.1.3	Stratospheric aerosol	271
9.2	Microphysics of cloud	272
9.2.1	Droplet growth by condensation	272
9.2.2	Droplet growth by collision	281
9.2.3	Growth of ice particles	282
9.3	Macroscopic characteristics of cloud	284
9.3.1	Formation and classification of cloud	284
9.3.2	Microphysical properties of cloud	292
9.3.3	Cloud dissipation	293
9.3.4	Cumulus detrainment: Influence on the environment	294
9.4	Radiative transfer in aerosol and cloud	298
9.4.1	Scattering by molecules and particles	298
9.4.2	Radiative transfer in a cloudy atmosphere	305
9.5	Roles of cloud and aerosol in climate	315
9.5.1	Involvement in the global energy budget	315
9.5.2	Involvement in chemical processes	324
	Suggested references	324
	Problems	325
10	Atmospheric motion	328
10.1	Description of atmospheric motion	329
10.2	Kinematics of fluid motion	330

10.3	The material derivative	334
10.4	Reynolds' transport theorem	334
10.5	Conservation of mass	336
10.6	The momentum budget	337
10.6.1	Cauchy's Equations of Motion	337
10.6.2	Momentum equations in a rotating reference frame	339
10.7	The first law of thermodynamics	341
	Suggested references	343
	Problems	343
11	Atmospheric equations of motion	345
11.1	Curvilinear coordinates	345
11.2	Spherical coordinates	348
11.2.1	The traditional approximation	353
11.3	Special forms of motion	354
11.4	Prevailing balances	355
11.4.1	Motion-related stratification	355
11.4.2	Scale analysis	356
11.5	Thermodynamic coordinates	358
11.5.1	Isobaric coordinates	358
11.5.2	Log-pressure coordinates	363
11.5.3	Iisentropic coordinates	365
	Suggested references	369
	Problems	369
12	Large-scale motion	371
12.1	Geostrophic equilibrium	371
12.1.1	Motion on an f plane	374
12.2	Vertical shear of the geostrophic wind	376
12.2.1	Classes of stratification	377
12.2.2	Thermal wind balance	379
12.3	Frictional geostrophic motion	380
12.4	Curvilinear motion	382
12.4.1	Inertial motion	384
12.4.2	Cyclostrophic motion	385
12.4.3	Gradient motion	385
12.5	Weakly divergent motion	386
12.5.1	Barotropic nondivergent motion	386
12.5.2	Vorticity budget under baroclinic stratification	387
12.5.3	Quasi-geostrophic motion	392
	Suggested references	395
	Problems	395
13	The planetary boundary layer	398
13.1	Description of turbulence	399
13.1.1	Reynolds decomposition	401
13.1.2	Turbulent diffusion	403

16 Dynamic stability	515
16.1 Inertial instability	515
16.2 Shear instability	517
16.2.1 Necessary conditions for instability	517
16.2.2 Barotropic and baroclinic instability	519
16.3 The Eady model	520
16.4 Nonlinear considerations	526
Suggested references	531
Problems	531
17 Influence of the ocean	533
17.1 Composition and structure	533
17.1.1 Stratification	536
17.1.2 Motion	538
17.2 Role in the heat budget	541
17.3 Role in the carbon cycle	544
17.4 The wind-driven circulation	546
17.4.1 The Ekman layer	546
17.4.2 Sverdrup balance	551
17.5 The buoyancy-driven circulation	553
17.6 Interannual changes	556
Suggested references	564
Problems	564
18 Interaction with the stratosphere	566
18.1 Ozone photochemistry	567
18.1.1 The chemical family	567
18.1.2 Photochemical equilibrium	568
18.2 Involvement of other species	570
18.2.1 Nitrous oxide	571
18.2.2 Chlorofluorocarbons	572
18.2.3 Methane	574
18.3 Motion	575
18.3.1 The Brewer-Dobson circulation	576
18.3.2 Wave driving of mean meridional motion	577
18.3.3 Transformed Eulerian description	581
18.4 Sudden stratospheric warmings	585
18.5 The quasi-biennial oscillation	589
18.6 Direct interaction with the troposphere	591
18.7 Heterogeneous chemical reactions	596
18.8 Interannual changes	602
Suggested references	615
Problems	615
Appendix A: Conversion to SI units	617
Appendix B: Thermodynamic properties of air and water	618
Appendix C: Physical constants	619

Appendix D: Vector identities	620
Appendix E: Curvilinear coordinates	621
Appendix F: Pseudo-adiabatic chart	623
Appendix G: Acronyms	625
Answers to selected problems	627
References	633
Index	647

Preface

Global measurements from space, coupled with large-scale computer models, have widened the perspective of atmospheric science along with its subdiscipline, the study of climate. Supporting those tools are proxy records of previous climate upon which rest interpretations of the current state of the Earth-atmosphere system. While opening new avenues of investigation, these modern tools have introduced increasingly complex questions. Many concern the tools themselves. Uncertainties surround the interpretation of observations, especially proxy records of previous climate, how key physical processes are represented in Global Climate Models (GCMs), and discrepancies between those models. These uncertainties make an understanding of the controlling physical processes and limitations that surround their description essential for drawing reliable insight into the Earth-atmosphere system.

Emerging simultaneously with technological advances has been growing concern over the role of humans in global climate. Buttressed by wide-ranging claims of environmental consequences, such concern has been pushed into the limelight of major national and international policy. The popular ascent of climate research has not been without criticism. Notable are concerns over rigor and critical analysis, whereby (i) proxies of previous climate, relied upon in interpretations of current climate, are often remote and ambiguous and (ii) insight into underlying physical mechanisms has been supplanted by models which, although increasingly complex, remain, in many respects, primitive and poorly understood. Despite technological advances in observing the Earth-atmosphere system and in computing power, strides in predicting its evolution reliably – on climatic time scales and with regional detail – have been limited. The pace of progress reflects the interdisciplinary demands of the subject. Reliable simulation, adequate to reproduce the observed record of climate variation, requires a grasp of mechanisms from different disciplines and of how those mechanisms are interwoven in the Earth-atmosphere system.

Historically, students of the atmosphere and climate have had proficiency in one of the physical disciplines that underpin the subject, but not in the others. Under the fashionable umbrella of climate science, many today do not have proficiency in even

one. What is today labeled climate science includes everything from archeology of the Earth to superficial statistics and a spate of social issues. Yet, many who embrace the label have little more than a veneer of insight into the physical processes that actually control the Earth-atmosphere system, let alone what is necessary to simulate its evolution reliably. Without such insight and its application to resolve major uncertainties, genuine progress is unlikely.

The atmosphere is the heart of the climate system, driven through interaction with the sun, continents, and ocean. It is the one component that is comprehensively observed. For this reason, the atmosphere is the central feature against which climate simulations must ultimately be validated.

This book builds on a forerunner, *Fundamentals of Atmospheric Physics*. It has been expanded to include climate, while streamlining the presentation for undergraduates in science, mathematics, and engineering. Advanced material, suitable as a resource for graduate students and researchers, has been retained (distinguished by shading). The treatment focuses upon physical concepts, which are developed from first principles. It integrates five major themes:

1. Atmospheric Thermodynamics;
2. Hydrostatic Equilibrium and Stability;
3. Radiation, Cloud, and Aerosol;
4. Atmospheric Dynamics and the General Circulation;
5. Interaction with the Ocean and Stratosphere.

Cornerstones of modern research, these themes are developed in a balance of theory and applications. Each is illustrated with manifestations on an individual day, the same day used to illustrate other themes. In this fashion, the Earth-atmosphere system is dissected in contemporaneous properties, revealing interactions among them. Supporting the development are detailed solutions to selected problems.

Chapter 1 presents an overview of the Earth-atmosphere system, describing its composition, structure, and energetics. It culminates in a discussion of global mean temperature, its relationship to atmospheric composition, and issues surrounding uncertainties in instrumental and proxy records of climate. Chapters 2–5 are devoted to atmospheric thermodynamics. Developed from a Lagrangian perspective, the discussion concentrates on heterogeneous systems that figure in considerations of cloud and its interaction with radiation, as well as the role of water vapor in the global energy budget. Hydrostatic equilibrium and stability are treated in Chapters 6 and 7, which develop their roles in convection and its influence on thermal and humidity structure. Chapters 8 and 9 focus on atmospheric radiation, cloud, and aerosol. After developing the laws governing radiative transfer, the presentation moves to the energetics of radiative and radiative-convective equilibrium. It then considers climate feedback mechanisms, which are discussed in relation to major contributors to the greenhouse effect, and their simulation in GCMs. Chapters 10–16 are devoted to atmospheric dynamics and the general circulation. The perspective is then transformed, via Reynolds' transport theorem, to the Eulerian description of behavior. Large-scale motion is first treated in terms of geostrophic and hydrostatic equilibrium and then extended to vorticity dynamics and quasi-geostrophic motion. The general circulation is motivated by a zonally symmetric model of heat transfer, setting the stage for baroclinic instability. Supporting it is a treatment of thermal properties of the Earth's surface, persistent

features of the circulation, and interannual fluctuations that comprise climate variability. The presentation then turns in Chapter 17 to the ocean, its structure, dynamics, and how it influences the atmosphere. The book closes with a treatment of the stratosphere, issues surrounding ozone, and interactions with the troposphere.

This book has benefited from interaction with numerous colleagues and students. In addition to those received earlier, contributions and feedback were generously provided by W. Bourke, J. Frederiksen, R. Madden, E. Titova, D. Toohey, and J. Wu. Figures were skillfully prepared by J. Davis and D. Oliver. Lastly, I am grateful for the understanding and encouragement of my son, without which this book would not have been completed.

Murry L. Salby

Prelude

The most fruitful areas for growth of the sciences are those between established fields. Science has been increasingly the task of specialists, in fields which show a tendency to grow progressively narrower. Important work is delayed by the unavailability in one field of results that may have already become classical in the next field. It is these boundary regions of science that offer the richest opportunities to the qualified investigator.

Norbert Wiener

The Earth-atmosphere system

1.1 INTRODUCTION

The Earth's atmosphere is the gaseous envelope surrounding the planet. Like other planetary atmospheres, it figures centrally in transfers of energy between the sun, the Earth, and deep space. It also figures in transfers of energy from one region of the globe to another. By maintaining thermal equilibrium, such transfers determine the Earth's climate. However, among neighboring planets, the Earth's atmosphere is unique because it is related closely to ocean and surface processes that, together with the atmosphere, form the basis for life.

Because it is a fluid system, the atmosphere is capable of supporting a wide spectrum of motions. These range from turbulent eddies of a few meters to circulations with dimensions of the Earth itself. By rearranging mass, air motion influences other atmospheric components such as water vapor, ozone, and cloud, which figure prominently in radiative and chemical processes. Such influence makes the atmospheric circulation a key ingredient of the global energy budget.

1.1.1 Descriptions of atmospheric behavior

The mobility of a fluid system makes its description complex. Atmospheric motion redistributes mass and constituents into a variety of complex configurations. Like any fluid system, the atmosphere is governed by the laws of continuum mechanics. They can be derived from the laws of mechanics and thermodynamics that govern a discrete fluid body by generalizing those laws to a continuum of such systems. In the atmosphere, the discrete system to which these laws apply is an infinitesimal fluid element, or *air parcel*, which is defined by a fixed collection of matter.

Two frameworks are used to describe atmospheric behavior. The *Eulerian description* represents atmospheric behavior in terms of field properties, such as the instantaneous distributions of temperature, motion, and constituents. Governed by partial differential equations, the field description of atmospheric behavior is convenient for numerical applications. The *Lagrangian description* represents atmospheric behavior in terms of the properties of individual air parcels (e.g., in terms of their instantaneous positions, temperatures, and constituent concentrations). Because it focuses on transformations of properties within an air parcel and on interactions between that system and its environment, the Lagrangian description offers conceptual as well as certain diagnostic advantages. For this reason, the basic principles governing atmospheric behavior are developed in this text from a Lagrangian perspective.

In the Lagrangian framework, the system considered is an individual air parcel moving through the circulation. Although it may change in form through deformation and in composition through thermodynamic and chemical transformations, this system is uniquely identified by the matter comprising it initially. Mass can be transferred across the boundary of an air parcel through molecular diffusion and turbulent mixing. However, such transfers are slow enough to be ignored for many applications. An individual parcel can then change only through interaction with its environment and through internal transformations that alter its composition and state.

1.1.2 Mechanisms influencing atmospheric behavior

Of the factors influencing atmospheric behavior, gravity is the single most important. Even though it has no upper boundary, the atmosphere is contained by the gravitational field of the Earth, which prevents atmospheric mass from escaping to space. Because it is such a strong body force, gravity determines many atmospheric properties. Most immediate is the geometry of the atmosphere. Atmospheric mass is concentrated in the lowest 10 km - less than 1% of the Earth's radius. Gravitational attraction has compressed the atmosphere into a shallow layer above the Earth's surface in which mass and constituents are stratified vertically: They are layered.

Through stratification of mass, gravity imposes a strong kinematic constraint on atmospheric motion. Circulations with dimensions greater than a few tens of kilometers are quasi-horizontal. Vertical displacements of air are then much smaller than horizontal displacements. Under these circumstances, constituents such as water vapor and ozone fan out in layers or "strata." Vertical displacements are comparable to horizontal displacements only in small-scale circulations such as convective cells and fronts, which have horizontal dimensions comparable to the vertical scale of the mass distribution.

The compressibility of air complicates the description of atmospheric behavior by enabling the volume of a parcel to change as it experiences changes in surrounding pressure. Therefore, concentrations of mass and constituents for the parcel can change, even though the number of molecules remains fixed. The concentration of a chemical constituent can also change through internal transformations, which alter the number of a particular type of molecule. For example, condensation decreases the abundance of water vapor in an air parcel that passes through a cloud system. Photodissociation of O_2 will increase the abundance of ozone in a parcel that passes through a region of sunlight.

Exchanges of energy with its environment and transformations between one form of energy and another likewise alter the properties of an air parcel. By expanding, an

air parcel exchanges energy mechanically with its environment through work that it performs on the surroundings. Heat transfer, as occurs through absorption of radiant energy and conduction with the Earth's surface, represents a thermal exchange of energy with a parcel's environment. Absorption of water vapor by an air parcel (e.g., through contact with a warm ocean surface) has a similar effect. When the vapor condenses, latent heat of vaporization carried by the vapor is released to the surrounding molecules of dry air that comprise the parcel. If the condensed water then precipitates back to the Earth's surface, this process leads to a net transfer of heat from the Earth's surface to the parcel.

The Earth's rotation, like gravity, exerts an important influence on atmospheric motion and, hence, on distributions of atmospheric properties. Because the Earth is a noninertial reference frame, the conventional laws of mechanics do not hold; they must be modified to account for its acceleration. Apparent forces introduced by the Earth's rotation are responsible for properties of the large-scale circulation, in particular, the flow of air around centers of low and high pressure. Those forces also inhibit meridional, e.g., NS (North-South) motion. Consequently, they inhibit transfers of heat and constituents between the equator and poles. For this reason, rotation tends to stratify properties meridionally, just as gravity tends to stratify them vertically.

The physical processes described in the preceding paragraphs do not operate independently. They are interwoven in a complex fabric of radiation, chemistry, and dynamics that govern the Earth-atmosphere system. Interactions among these can be just as important as the individual processes themselves. For instance, radiative transfer controls the thermal structure of the atmosphere, which determines the circulation. Transport by the circulation, in turn, influences the distributions of radiatively active components such as water vapor, ozone, and cloud. In view of their interdependence, understanding how one of these processes influences behavior requires an understanding of how that process interacts with others. This feature makes the study of the Earth-atmosphere system an eclectic one, involving the integration of many different physical principles. This book develops the most fundamental of these.

1.2 COMPOSITION AND STRUCTURE

The Earth's atmosphere consists of a mixture of gases, mostly molecular nitrogen (78% by volume) and molecular oxygen (21% by volume); see Table 1.1. Water vapor, carbon dioxide, and ozone, along with other minor constituents, comprise the remaining 1% of the atmosphere. Although present in very small abundances, trace species such as water vapor and ozone play a key role in the energy balance of the Earth through their involvement in radiative processes. Because they are created and destroyed in particular regions and are linked to the circulation through transport, these and other minor species are highly variable. For this reason, trace species are treated separately from the primary atmospheric constituents, which are referred to simply as "dry air."

1.2.1 Description of air

The starting point for describing atmospheric behavior is the ideal gas law

$$\begin{aligned} pV &= nR^*T \\ &= \frac{m}{M}R^*T \\ &= mRT, \end{aligned} \tag{1.1}$$

Table 1.1. Atmospheric Composition. Constituents are listed with volume mixing ratios representative of the Troposphere or Stratosphere, how the latter are distributed vertically, and controlling processes

Constituent	Tropospheric Mixing Ratio	Vertical Distribution (Mixing Ratio)	Controlling Processes
N_2	.7808	Homogeneous	Vertical Mixing
O_2	.2095	Homogeneous	Vertical Mixing
* H_2O	≤ 0.030	Decreases sharply in Troposphere Increases in Stratosphere Highly Variable	Evaporation, Condensation, Transport Production by CH_4 Oxidation
Ar	.0093	Homogeneous	Vertical Mixing
* CO_2	380 ppmv	Homogeneous	Vertical Mixing Production by Surface and Anthropogenic Processes
* O_3	10 ppmv [§]	Increases sharply in Stratosphere Highly Variable	Photochemical Production in Stratosphere; secondarily through pollution in troposphere Destruction at Surface Transport
* CH_4	1.8 ppmv	Homogeneous in Troposphere Decreases in Middle Atmosphere	Production by Surface Processes Oxidation Produces H_2O
* N_2O	320 ppbv	Homogeneous in Troposphere Decreases in Middle Atmosphere	Production by Surface and Anthropogenic Processes Dissociation in Middle Atmosphere Produces NO Transport
* CO	70 ppbv	Decreases in Troposphere Increases in Stratosphere	Production Anthropogenically and by Oxidation of CH_4 Transport
NO	0.1 ppbv [§]	Increases Vertically	Production by Dissociation of N_2O Catalytic Destruction of O_3
* $CFC-11$	0.2 ppbv	Homogeneous in Troposphere	Industrial Production
* $CFC-12$	0.5 ppbv	Decreases in Stratosphere	Mixing in Troposphere
* $HFC-134A$	30 ppt		Photo-dissociation in Stratosphere

* Radiatively active

[§] Stratospheric value

which constitutes the equation of state for a pure (single-component) gas. In (1.1), p , T , and M denote the pressure, temperature, and molar weight of the gas, and V , m , and $n = \frac{m}{M}$ refer to the volume, mass, and molar abundance of a fixed collection of matter (e.g., an air parcel). The *specific gas constant* R is related to the *universal gas constant* R^* through

$$R = \frac{R^*}{M}. \quad (1.2)$$

Equivalent forms of the ideal gas law that do not depend on the dimension of the system are

$$\begin{aligned} p &= \rho RT \\ pv &= RT, \end{aligned} \quad (1.3)$$

where ρ and $v = \frac{1}{\rho}$ (also denoted α) are the density and *specific volume* of the gas.

Because it is a mixture of gases, air obeys similar relationships. So do its individual components. The *partial pressure* p_i of the i th component is that pressure the i th component would exert in isolation at the same volume and temperature as the mixture. It satisfies the equation of state

$$p_i V = m_i R_i T, \quad (1.4.1)$$

where R_i is the specific gas constant of the i th component. Similarly, the *partial volume* V_i is that volume the i th component would occupy in isolation at the same pressure and temperature as the mixture. It satisfies the equation of state

$$p V_i = m_i R_i T. \quad (1.4.2)$$

Dalton's law asserts that the pressure of a mixture of gases equals the sum of their partial pressures

$$p = \sum_i p_i. \quad (1.5)$$

Likewise, the volume of the mixture equals the sum of the partial volumes¹

$$V = \sum_i V_i. \quad (1.6)$$

The equation of state for the mixture can be obtained by summing (1.4) over all of the components

$$pV = T \sum_i m_i R_i.$$

Then, defining the mean specific gas constant

$$\bar{R} = \frac{\sum_i m_i R_i}{m} \quad (1.7)$$

yields the equation of state for the mixture

$$pV = m \bar{R} T. \quad (1.8)$$

The mean molar weight of the mixture is defined by

$$\bar{M} = \frac{m}{n}. \quad (1.9)$$

Because the molar abundance of the mixture is equal to the sum of the molar abundances of the individual components,

$$n = \sum_i \frac{m_i}{M_i},$$

(1.9) may be expressed

$$\bar{M} = \frac{R^* m}{\sum_i m_i \left(\frac{R^*}{M_i} \right)}.$$

¹ These are among several consequences of the Gibbs-Dalton law, which relates the properties of a mixture to properties of the individual components (e.g., Keenan, 1970).

Then applying (1.2) for the i th component together with (1.7) leads to

$$\bar{R} = \frac{R^*}{M}. \quad (1.10)$$

Equation (1.10) is analogous to (1.2) for a single-component gas.

Because of their involvement in radiative and chemical processes, variable components of air must be quantified. The “absolute concentration” of the i th species is measured by its density ρ_i or, alternatively, by its *number density*

$$[i] = \left(\frac{N_A}{M_i} \right) \rho_i \quad (1.11)$$

(also denoted n_i), where N_A is Avogadro’s number and M_i is the molar weight of the species. Partial pressure p_i and partial volume V_i are other measures of absolute concentration.

The compressibility of air makes absolute concentration an ambiguous measure of a constituent’s abundance. Even if a constituent is passive, namely if the number of molecules inside an individual parcel is fixed, its absolute concentration can change through changes of volume. For this reason, a constituent’s abundance is more faithfully described by the “relative concentration,” which is referenced to the overall abundance of air or simply dry air. The relative concentration of the i th species is measured by the *molar fraction*

$$N_i = \frac{n_i}{n}. \quad (1.12)$$

Dividing (1.4) by (1.8) and applying (1.2) for the i th component leads to

$$N_i = \frac{p_i}{p} = \frac{V_i}{V}. \quad (1.13)$$

Molar fraction uses as a reference the molar abundance of the mixture, which can vary through changes of individual species. A more convenient measure of relative concentration is mixing ratio. The *mass mixing ratio* of the i th species

$$r_i = \frac{m_i}{m_d}, \quad (1.14)$$

where the subscript d refers to dry air, is dimensionless. It is expressed in g kg^{-1} for tropospheric water vapor and in parts per million by mass (ppmm, or simply ppm) for stratospheric ozone. Unlike molar abundance, the reference mass m_d is constant for an individual air parcel. If the i th species is passive, namely if it does not undergo a transformation of phase or a chemical reaction, its mass m_i is also constant. The mixing ratio r_i is then fixed for an individual air parcel.

For a trace species, such as water vapor or ozone, the mixing ratio is closely related to the molar fraction

$$N_i \cong \frac{r_i}{\epsilon_i}, \quad (1.15.1)$$

where

$$\epsilon_i = \frac{M_i}{M_a}, \quad (1.15.2)$$

because the mass of air in the presence of such species is nearly identical to that of dry air. The *volume mixing ratio* provides similar information. It is distinguished from mass mixing ratio by dimensions such as parts per million by volume (ppmv) for stratospheric ozone (Probs. 1.2, 1.3). From (1.13) and (1.12), it follows that the volume mixing ratio is approximately equal to the molar fraction. Each measures the relative abundance of molecules of the i th species.

As noted, the mixing ratio of a passive species is fixed for an individual air parcel. A property that is invariant for individual parcels is said to be *conserved*. Although constant for individual air parcels, a conserved property is generally not constant in space and time. Unless that property happens to be homogeneous, its distribution must vary spatially and temporally, as parcels with different values exchange positions. A conserved property is a *material tracer* because particular values track the motion of individual air parcels. Thus, tracking particular values of r_i provides a description of how air is rearranged by the circulation and, therefore, of how all conserved species are redistributed.

1.2.2 Stratification of mass

By confining mass to a shallow layer above the Earth's surface, gravity exerts a profound influence on atmospheric behavior. If vertical accelerations are ignored, then Newton's second law of motion applied to the column of air between some level at pressure p and a level incrementally higher at pressure $p + dp$ (Fig. 1.1) reduces to a balance between the weight of that column and the net pressure force acting on it

$$pdA - (p + dp)dA = \rho g dV,$$

where g denotes the acceleration of gravity, or

$$\frac{dp}{dz} = -\rho g. \quad (1.16)$$

This simple form of mechanical equilibrium is known as *hydrostatic balance*. It is a good approximation even if the atmosphere is in motion because, for large-scale circulations, vertical displacements of air are small. This feature renders vertical acceleration two to three orders of magnitude smaller than the forces in (1.16). Applying the same analysis between the pressure p and the top of the atmosphere (where p vanishes) illustrates the origin of atmospheric pressure: The pressure at any level must equal the weight of the atmospheric column of unit cross-sectional area above that level.

Owing to the compressibility of air, the density in (1.16) is not constant. It depends on the air's pressure through the gas law. Eliminating ρ with (1.3) and integrating from the surface to an altitude z yields

$$\frac{p}{p_s} = e^{-\int_{z_s}^z \frac{dz'}{H(z')}}, \quad (1.17.1)$$

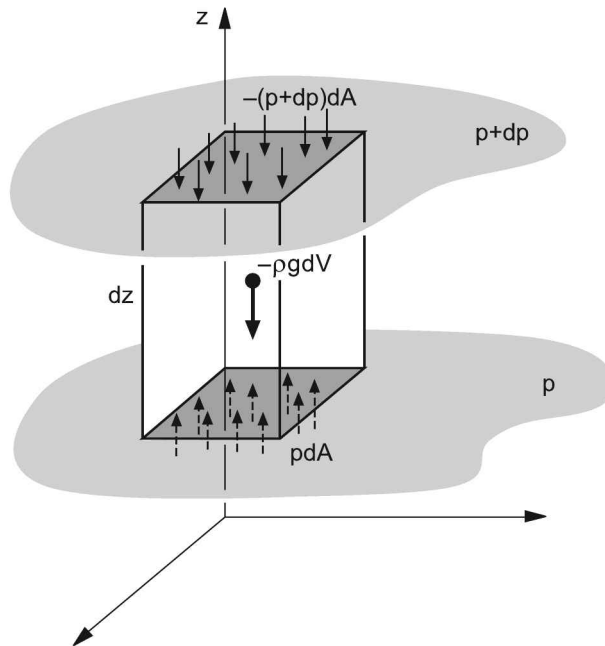


Figure 1.1 Hydrostatic balance for an incremental atmospheric column of cross-sectional area dA and height dz , bounded vertically by isobaric surfaces at pressures p and $p + dp$.

where

$$H(z) = \frac{RT(z)}{g} \quad (1.17.2)$$

is the pressure *scale height* and p_s is the surface pressure. The scale height represents the characteristic vertical dimension of the mass distribution. A function of altitude, it varies from about 8 km near the Earth's surface to 6 km in very cold regions of the atmosphere.

As illustrated by Fig. 1.2, global-mean pressure and density decrease with altitude approximately exponentially. Pressure decreases from about 1000 hPa or 10^5 Pascals (Pa) at the surface to only 10% of that value at an altitude of 15 km (two scale heights).² According to hydrostatic balance, 90% of the atmosphere's mass then lies beneath this level. Pressure decreases by another factor of 10 for each additional 15 km of altitude. Density decreases with altitude at about the same rate, from a surface value of about 1.2 kg m^{-3} . The sharp upward decrease of pressure implies that isobaric surfaces, along which $p = \text{const}$, are quasi-horizontal. Deflections of those surfaces introduce comparatively small horizontal variations of pressure that drive atmospheric motion.

² The historical unit of pressure, millibar (mb), has been replaced by its equivalent in the Standard International (SI) system of units, the hectoPascal (hPa), where $1 \text{ hPa} = 100 \text{ Pa} = 1 \text{ mb}$. See Appendix A for conversions between the SI system of units and others.

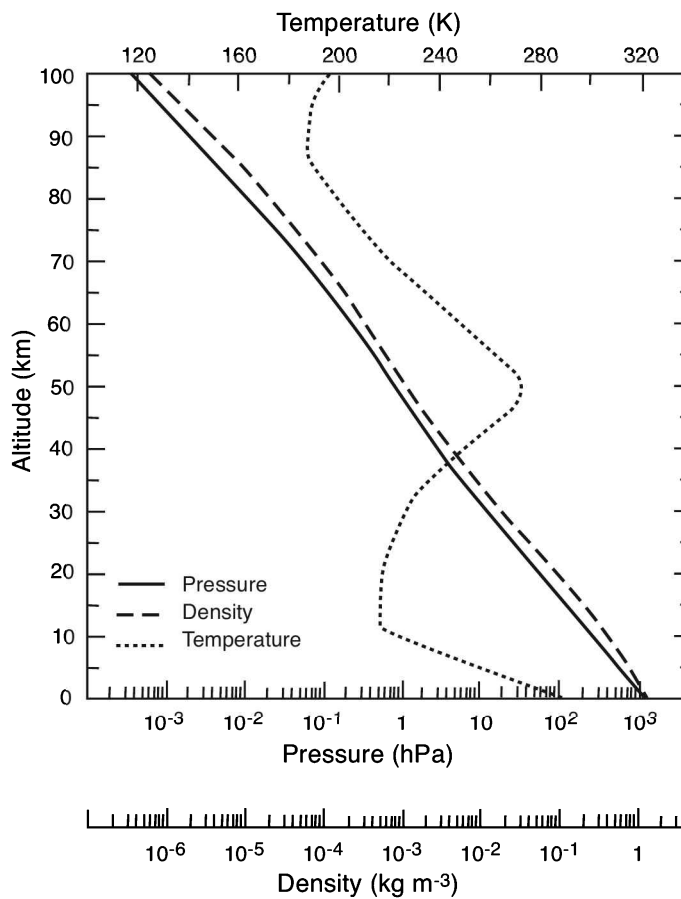


Figure 1.2 Global-mean pressure (solid), density (dashed), and temperature (dotted), as functions of altitude. *Source: U.S. Standard Atmosphere (1976).*

Above 100 km, pressure and density also decrease exponentially (Fig. 1.3), but at a rate which differs from that below and which varies gradually with altitude. The distinct change of behavior near 100 km marks a transition in the processes that control the stratification of mass and the composition of air. The mean free path of molecules is determined by the frequency of collisions. It varies inversely with air density. Consequently, the mean free path increases exponentially with altitude, from about 10^{-7} m at the surface to of order 1 m at 100 km. Because it controls molecular diffusion, the mean free path determines properties of air such as viscosity and thermal conductivity. Diffusion of momentum and heat supported by those properties dissipate atmospheric motion by destroying gradients of velocity and temperature.

Below 100 km, the mean free path is short enough for turbulent eddies in the circulation to be only weakly damped by molecular diffusion. At those altitudes, bulk transport by turbulent air motion dominates diffusive transport of atmospheric constituents. Turbulence stirs different gases with equal efficiency. Mixing ratios of

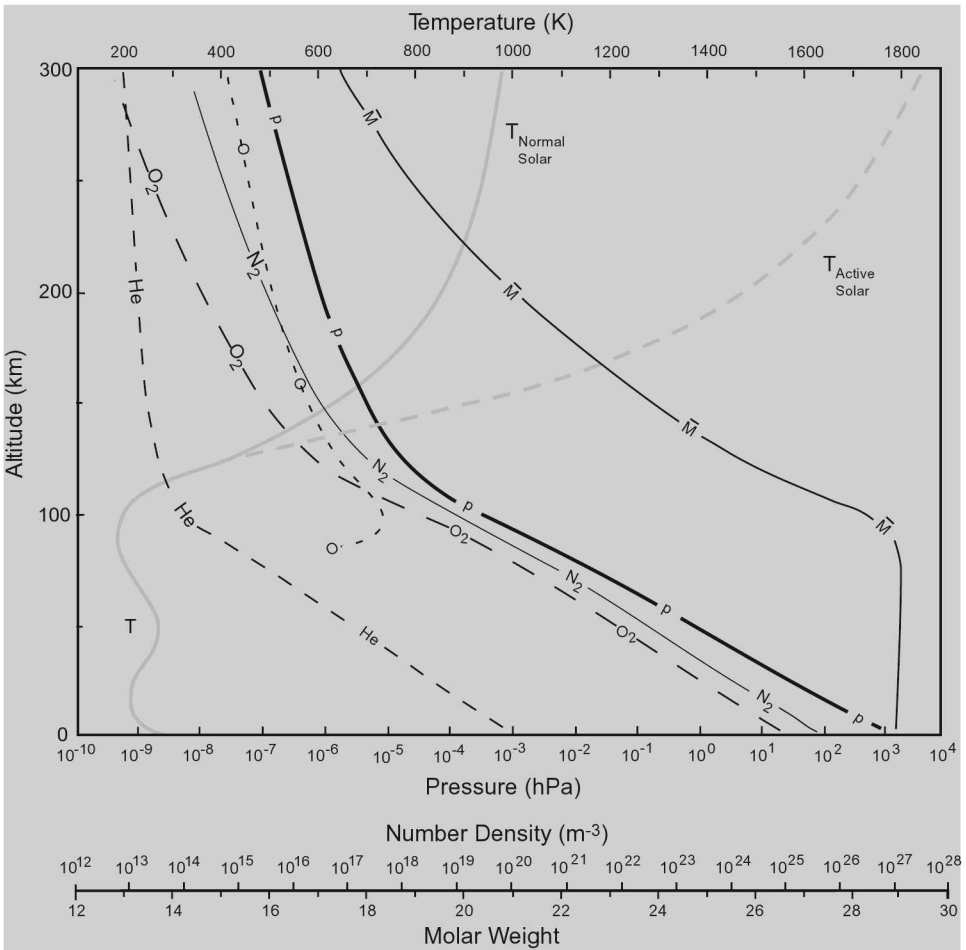


Figure 1.3 Global-mean pressure (bold), temperature (stippled), mean molar weight (solid), and number densities of atmospheric constituents as functions of altitude. *Source: U.S. Standard Atmosphere (1976).*

passive constituents are therefore homogeneous in this region. Those constituents are said to be “well mixed.” The densities of passive constituents then all decrease with altitude at the same exponential rate. This gives air a homogeneous composition, with constant mixing ratios $r_{N_2} \cong 0.78$, $r_{O_2} \cong 0.21$, and the constant gas properties³

$$M_d = 28.96 \text{ g mol}^{-1} \quad (1.18.1)$$

$$R_d = 287.05 \text{ J kg}^{-1} \text{ K}^{-1}. \quad (1.18.2)$$

The well-mixed region below 100 km is known as the *homosphere*.

³ Properties of dry air are tabulated in Appendix B, along with other thermodynamic constants.

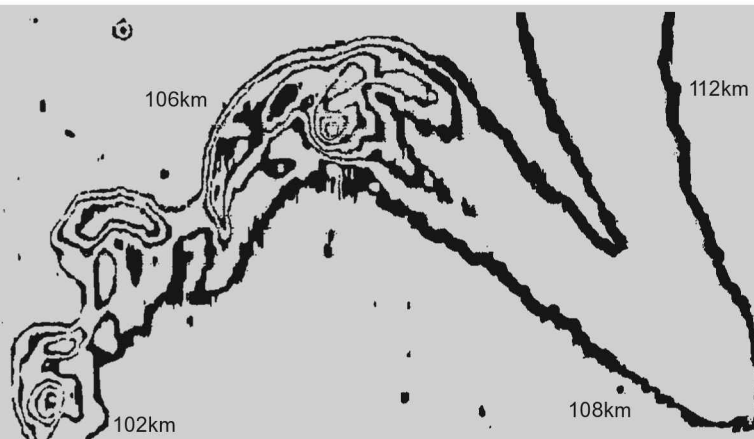


Figure 1.4 Constant-density contours of a chemical vapor trail released by a rocket traversing the turbopause. Beneath 107 km, the vapor trail is distorted by an array of turbulent eddies that form in the wake of the rocket. Above 107 km, the vapor trail remains laminar, reflecting the absence of turbulence, and expands under the action of molecular diffusion. Adapted from Roper (1977).

Above 100 km, the mean free path quickly becomes larger than turbulent displacements of air. As a result, turbulent air motion there is strongly damped by diffusion of momentum and heat. Diffusive transport then becomes the dominant mechanism for transferring properties vertically. The transition from turbulent transport to diffusive transport occurs at the *homopause* (also known as the *turbopause*), which has an average altitude of about 100 km. Figure 1.4 shows a rocket vapor trail traversing this region. Below 100 km, the trail is marked by turbulent eddies. Produced in the wake of the rocket, they homogenize different constituents with equal efficiency. Those eddies are conspicuously absent above 107 km. At this altitude, molecular diffusion suppresses turbulent air motion, becoming the prevailing form of vertical transport.

The region above the homopause and below 500 km is known as the *heterosphere*. In the heterosphere, air flow is nearly laminar, dominated by molecular diffusion. Because it operates on gases according to their molar weights, molecular diffusion stratifies constituents so that the heaviest species O_2 decreases with altitude more rapidly than the second heaviest species N_2 and so forth (Fig. 1.3). For this reason, the composition of air changes with altitude in the heterosphere, as is evidenced by the mean molar weight in Fig. 1.3. Constant below the homopause, \bar{M} changes abruptly near 100 km, above which it decreases upward monotonically.

Diffusive separation is primarily responsible for the stratification of constituents in the heterosphere. However, photo-dissociation also plays a role. Energetic ultraviolet (UV) radiation incident at the top of the atmosphere dissociates molecular oxygen. This process provides an important source of atomic oxygen at these altitudes. In fact, not far above the homopause, O becomes the dominant form of oxygen.

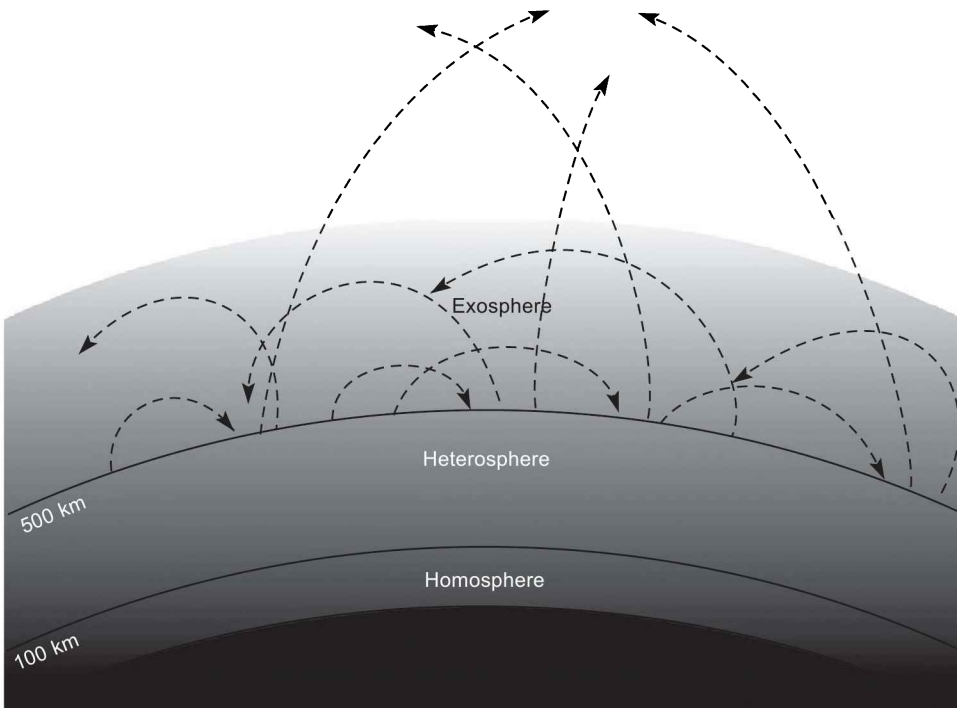


Figure 1.5 Schematic cross section of the atmosphere illustrating the homosphere, heterosphere, and the exosphere, in which molecular trajectories are shown.

Photo-dissociation of H_2O at lower altitudes by less energetic UV radiation liberates atomic hydrogen. H produced in this manner is gradually mixed and diffused to higher altitude. By dissociating such molecules, energetic radiation is filtered from the solar spectrum that penetrates to lower levels.

In the homosphere and heterosphere, atmospheric molecules interact strongly through frequent collisions. Their interaction becomes infrequent above an altitude of about 500 km, known as the *critical level*. At this altitude, molecular collisions are so rare that a significant fraction of the molecules pass out of the atmosphere without experiencing a single collision. Known as the *exosphere*, the region above the critical level contains molecules that leave the denser atmosphere and move out into space. As is illustrated in Fig. 1.5, these molecules follow ballistic trajectories that are determined by the molecular velocity at the critical level and the gravitational attraction of the Earth. Most of the molecules in the exosphere are captured by the Earth's gravitational potential. They return to the denser atmosphere along parabolic trajectories. However, some molecules have velocities great enough to escape the Earth's gravitational potential entirely. Those are lost to deep space.

The escape velocity v_e is determined by the kinetic energy adequate to liberate a molecule from the potential well of the Earth's gravitational field. That energy equals the work performed to displace a molecule from the critical level

to infinity

$$\frac{1}{2}mv_e^2 = \int_a^\infty mg_0 \left(\frac{a}{r}\right)^2 dr, \quad (1.19)$$

where a is the Earth's radius, g_0 is the gravitational attraction averaged over the surface of the Earth, $m \cdot g_0 \left(\frac{a}{r}\right)^2$ is the molecule's local weight, dr is its local displacement, and the difference between a and the radial distance to the critical level is negligible. The escape velocity follows from (1.19) as

$$v_e = \sqrt{2g_0 a}. \quad (1.20)$$

For Earth, v_e has a value of about 11 km s⁻¹. Independent of molecular weight, v_e is the same for all molecules. However, different molecules do not all have the same distribution of velocities. Energy is equipartitioned in a molecular ensemble (see, e.g., Lee, Sears, and Turcotte, 1973). Lighter molecules thus have faster velocities than heavier ones. For this reason, lighter atmospheric constituents escape to space more readily than do heavier constituents.

The critical level of the Earth's atmosphere lies near 500 km. The temperature at this altitude is about 1000 K under conditions of normal solar activity. However, it can reach 2000 K under disturbed solar conditions (Fig. 1.3). Figure 1.6 shows for O and H the Boltzmann distributions of a molecular ensemble (see, e.g., Lee, Sears, and Turcotte, 1973), as functions of molecular velocity:

$$\frac{dn}{n} = \frac{4}{\sqrt{\pi}} \frac{v^2}{v_0^3} e^{-\left(\frac{v}{v_0}\right)^2} dv, \quad (1.21)$$

where $\frac{dn}{n}$ represents the fractional number of molecules having velocities in the range $(v, v + dv)$,

$$v_0 = \sqrt{\frac{2kT}{m}} \quad (1.22)$$

is the most probable velocity, m is the mass of the molecules, and k is the Boltzmann constant. The fraction of molecules with velocities exceeding the escape velocity is

$$\begin{aligned} \frac{\Delta n_e}{n} &= \int_{v_e}^\infty \frac{4}{\sqrt{\pi}} \frac{v^2}{v_0^3} e^{-\left(\frac{v}{v_0}\right)^2} dv \\ &\cong \frac{2}{\sqrt{\pi}} \left(\frac{v_e}{v_0}\right) e^{-\left(\frac{v_e}{v_0}\right)^2} \end{aligned} \quad (1.23)$$

for $v_0 \ll v_e$.

For atomic oxygen, the most probable velocity is $v_0 = 1.02$ km s⁻¹. The fraction of O molecules with velocities faster than v_e is only about 10⁻⁴⁵. A lower bound on the time to deplete all O molecules initially at the critical level (e.g., neglecting production of O molecules locally by photo-dissociation and diffusive transport from below) is given by the mean time between collisions divided by the fraction of molecules moving upward with $v > v_e$. Near 500 km, the mean time between collisions is of order 10 s. Hence the time for all O molecules initially at the critical level to escape the Earth's gravitational field is greater than 10⁴⁶ s. This is far greater than the 4 billion years that the Earth has existed. Heavier species are captured by the Earth's gravitational field even more effectively.

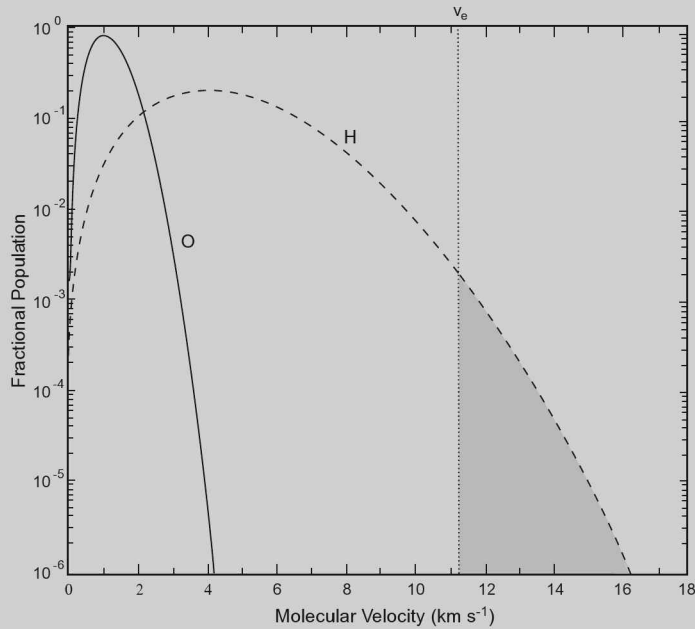


Figure 1.6 Boltzmann distribution of velocities for a molecular ensemble of oxygen atoms and hydrogen atoms. Escape velocity v_e for Earth also indicated.

The situation for hydrogen differs significantly. The population of H is distributed over much faster velocities, so many more molecules exceed the escape velocity v_e . The most probable velocity is 4.08 km s^{-1} , whereas the fraction of molecules with velocities faster than v_e is about 10^{-4} . By previous reasoning, a significant fraction of H molecules initially at the critical level is lost to deep space after only $10^5 \text{ s} \approx 1 \text{ day}$. Under conditions of disturbed solar activity, temperatures at the critical level are substantially greater. Hydrogen molecules are then boiled off of the atmosphere even faster. The rapid escape of atomic hydrogen from the Earth's gravitational field explains why H is found in the atmosphere only in very small abundances, despite its continual production by photo-dissociation of H_2O .

1.2.3 Thermal and dynamical structure

The atmosphere is categorized according to its thermal structure, which determines dynamical properties of individual regions. The simplest picture of atmospheric thermal structure is the vertical profile of global-mean temperature in Fig. 1.2. From the surface up to about 10 km, temperature decreases upward at a nearly constant *lapse rate*, which is defined as the rate at which temperature “decreases” with altitude. This layer immediately above the Earth's surface is known as the *troposphere*, which means “turning sphere.” Its title is symbolic of the convective overturning that characterizes this region. Having a global-mean lapse rate of about 6.5 K km^{-1} , the troposphere contains most of the behavior identified with weather. It is driven ultimately by surface heating, which is then transferred to the atmosphere. The upper boundary of the troposphere, or “tropopause,” lies at an altitude of about 10 km (100 hPa), marked by a minimum of temperature and a sharp change of lapse rate.

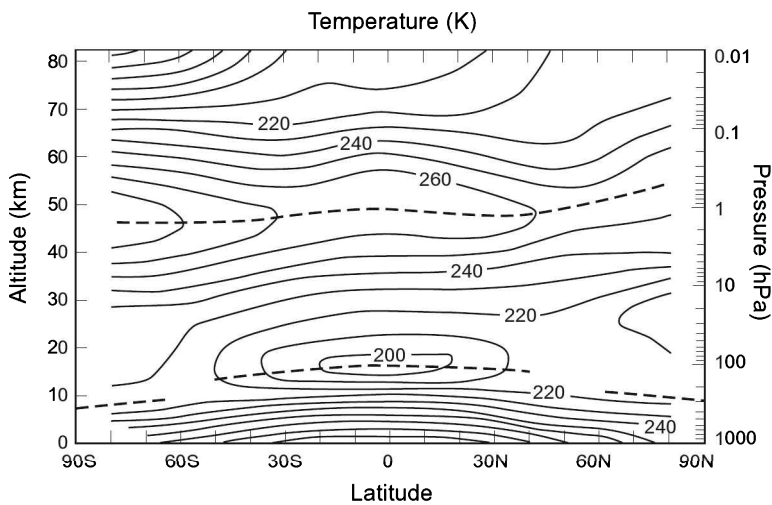


Figure 1.7 Zonal-mean temperature during northern winter as a function of latitude and altitude. Adapted from Fleming et al. (1988).

The region from the tropopause to an altitude of about 85 km is known as the *middle atmosphere*. Temperature first remains nearly constant and then increases upward in the *stratosphere*, which means “layered sphere.” Its title is symbolic of properties at these altitudes, which experience little vertical mixing. In the stratosphere, temperature increases upward (negative lapse rate). This thermal structure reflects ozone heating; the latter results from the absorption of solar UV by O_3 , which increases upward above the tropopause. Contrary to the troposphere, the stratosphere involves only weak vertical motions. It is dominated by radiative processes. The upper boundary of the stratosphere, or “stratopause,” lies at an altitude of about 50 km (1 hPa), where temperature reaches a maximum.

Above the stratopause, temperature again decreases with altitude in the *mesosphere*, where ozone heating diminishes. Convective motion and radiative processes are both important in the mesosphere. Meteor trails form in this region of the atmosphere; so do lower layers of the ionosphere during daylight hours. The “mesopause” lies at an altitude of about 85 km (.01 hPa), where a second minimum of temperature is reached.

Above the mesopause, temperature increases steadily in the *thermosphere* (cf. Fig. 1.3). Unlike lower regions, the thermosphere cannot be treated as an electrically neutral continuum. Ionization of molecules by energetic solar radiation produces a plasma of free electrons and ions. Oppositely charged, those particles interact differently with the Earth’s electric and magnetic fields. As is apparent in Fig. 1.3, this region of the atmosphere is influenced strongly by variations of solar activity. Below the mesopause, however, the influence of solar variations is limited.

A more complete picture of the thermal structure of the atmosphere is provided by the zonal-mean temperature \bar{T} , where $(\bar{\quad})$ denotes the longitudinal average. \bar{T} is plotted in Fig. 1.7, as a function of latitude and altitude, during northern winter. In the troposphere, temperature decreases with altitude and latitude. The tropopause, which is characterized by an abrupt change of lapse rate, is highest in the tropics (~16 km), where temperature has decreased to 200 K. It is lowest in polar regions

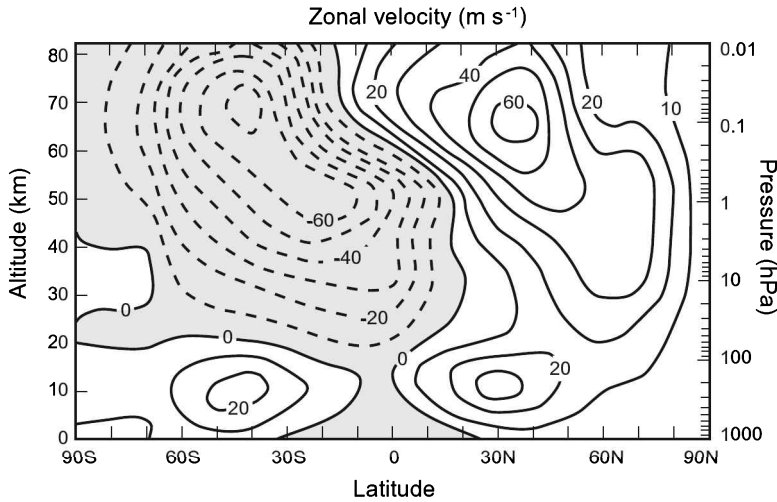


Figure 1.8 As in Fig. 1.7, but for the zonal component of velocity.

(~8 km), where temperature is 10–20 K warmer, even in perpetual darkness that prevails over the winter pole (*polar night*). A sharp change of zonal-mean lapse rate is not observed at mid-latitudes, symbolized by a break in the tropopause. Above the tropopause, temperature increases with altitude. It is warmest over the summer pole and decreases steadily to coldest temperature over the winter pole. In the mesosphere, where temperature again decreases with altitude, the horizontal temperature gradient is reversed. Temperature is actually coldest over the summer pole, which lies in perpetual sunlight (*polar day*). It increases steadily to warmest temperature over the winter pole, which lies in perpetual darkness. This peculiarity of the temperature distribution is incongruous with radiative considerations. It illustrates the importance of dynamics to establishing observed thermal structure.

The thermal structure in Fig. 1.7 is related closely to the zonal-mean circulation \bar{u} , which is shown in Fig. 1.8 at the same time of year. In the troposphere, the zonal (EW) velocity u is characterized by subtropical *jet streams*. They strengthen upward, maximizing near the tropopause. These jets describe circumpolar motion that is westerly in each hemisphere.⁴ Above the subtropical jets, zonal-mean wind first weakens with altitude. It then intensifies, but with opposite sign in the two hemispheres. In the winter hemisphere, westerlies intensify above the tropopause, forming the *polar-night jet*. There, wind speed reaches 60 m s⁻¹ in the lower mesosphere. In the summer hemisphere, westerlies weaken above the tropopause, replaced by easterlies that intensify up to the mesosphere. This easterly circulation attains speeds somewhat stronger than the westerly circulation in the winter hemisphere. At lower altitudes, it joins weak easterlies that prevail in the tropical troposphere.

On individual days, the circulation is more complex and variable than is represented in the zonally- and time-averaged distributions in Figs. 1.7 and 1.8. Figure 1.9a shows, for an individual day, the instantaneous circulation on the 500-hPa isobaric surface: that locus of points where the pressure equals 500 hPa. Contoured over the circulation is the height of the 500 hPa isobaric surface. It reflects the horizontal distribution of pressure.

⁴ In meteorological vernacular, *westerly* refers to motion from the west, whereas *easterly* refers to motion from the east.

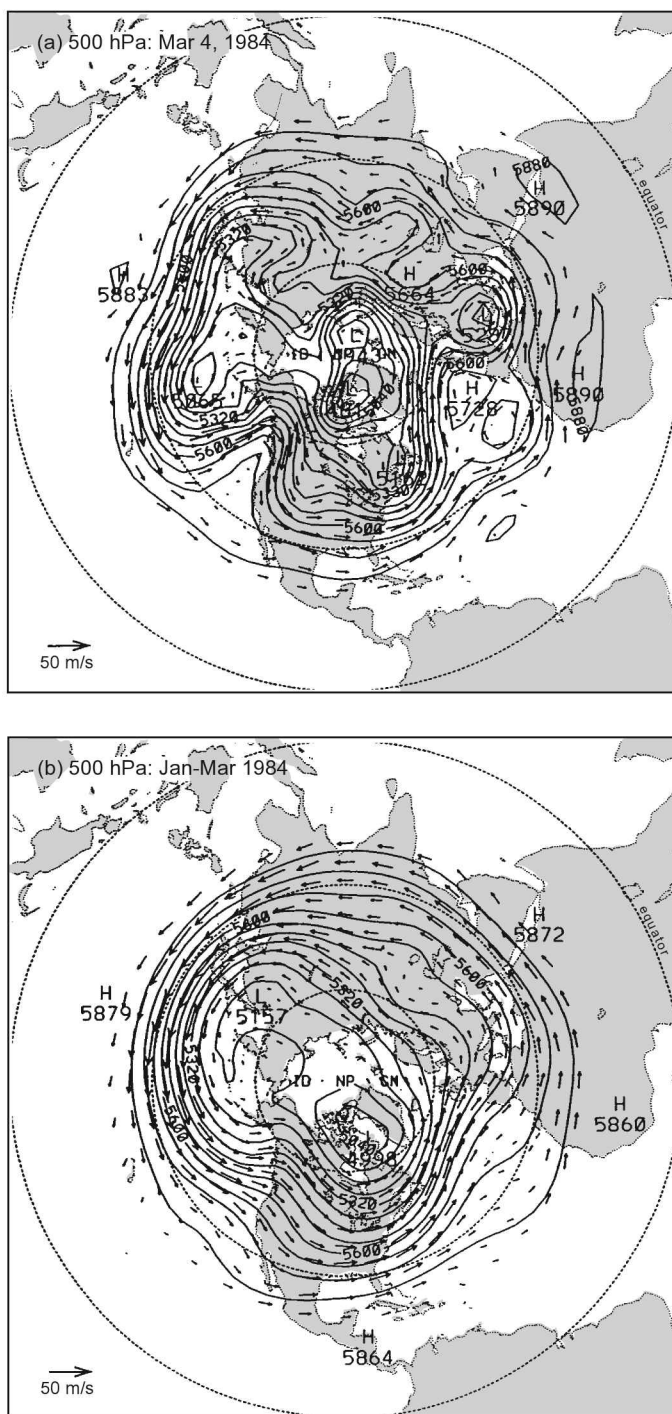


Figure 1.9 (a) Height (contours) of and horizontal velocity (vectors) on the 500-hPa isobaric surface for March 4, 1984. (b) Time-mean height of and velocity on the 500-hPa isobaric surface for January–March, 1984. Isobaric heights shown in meters. From National Meteorological Center (NMC) analyses.

Hydrostatic equilibrium (1.17) requires pressure and altitude to maintain a single-valued relationship (Fig. 1.2). Therefore, if one is known, so is the other. The altitude of the 500-hPa isobaric surface may then be interpreted analogously to the pressure on a surface of constant altitude $z \cong 5.5$ km, the mean height of the 500-hPa surface. Because pressure decreases upward monotonically, low altitude of the 500-hPa surface corresponds to low pressure on that surface of constant altitude (cf. Fig. 6.2). Similarly, high altitude of the 500-hPa surface corresponds to high pressure on that surface of constant altitude. Thus, centers of low altitude that punctuate the height of the 500-hPa surface in Fig. 1.9a imply centers of low pressure on the surface of constant altitude, $z \cong 5.5$ km. Centers of high altitude imply the reverse.

The 500-hPa surface slopes downward toward the pole, in the direction of decreasing tropospheric temperature (Figs. 1.9; 1.7). The circulation at 500 hPa is characterized by westerly circumpolar flow. It corresponds to the zonal-mean subtropical jet in Fig. 1.8. However, on the individual day shown, the jet stream is far from zonally symmetric. It is disturbed by half a dozen depressions of the 500-hPa surface. They are associated with *synoptic weather systems*. Those unsteady disturbances deflect the air stream meridionally (i.e., NS), distorting the circulation into a wavy pattern that meanders around the globe. Although highly disturbed, the circumpolar flow remains nearly tangential to contours of isobaric height.

In addition to synoptic weather systems, disturbances of global dimension also appear in the 500-hPa circulation. Known as *planetary waves*, these disturbances are manifested by a displacement of the circumpolar flow out of zonal symmetry, as is apparent at high latitude. They introduce a gradual undulation of the jet stream about latitude circles. Planetary waves are more evident in the time-mean circulation, shown in Fig. 1.9b. In it, unsteady synoptic weather systems have been filtered out – because they fluctuate on short time scales. The time-mean circulation exhibits more zonal symmetry. It therefore possesses greater correspondence to zonal-mean structure in Fig. 1.8 (e.g., the subtropical jet) than does the instantaneous circulation in Fig. 1.9a. Nonetheless, the time-mean flow is still disturbed, only on larger scales. Steady planetary waves are not removed by time averaging. They too displace the circumpolar flow out of zonal symmetry, deflecting the air stream across latitude circles. In addition, the time-mean circulation contains locally intensified jets. They mark the *North Atlantic storm track* and the *North Pacific storm track*. Accompanied by a steep meridional gradient of height, they are preferred sites of cyclone development. By determining these and other longitudinally dependent features of the time-mean circulation, planetary waves shape the climate of individual regions.

In the stratosphere, the circulation is also unsteady. However, the synoptic disturbances that prevail in the troposphere are not evident in the instantaneous circulation at 10 hPa, shown in Fig. 1.10a for the same day as in Fig. 1.9a. Instead, only planetary-scale disturbances appear at this level. The time-mean circulation at 10 hPa (Fig. 1.10b) is characterized by strong westerly flow. It corresponds to the polar-night jet in Fig. 1.8 and to a circumpolar vortex, the *polar-night vortex*. Like motion at 500 hPa, the time-mean circulation is still disturbed from zonal symmetry by steady planetary waves, which are not eliminated by averaging over time. By comparison, the instantaneous circulation in Fig. 1.10a is much more disturbed. Counterclockwise motion about the low, comprising the polar-night vortex, has been displaced off the pole, replaced by clockwise motion about a high that has amplified and temporarily invaded the polar cap. This disturbance deflects the air stream meridionally, which therefore undergoes

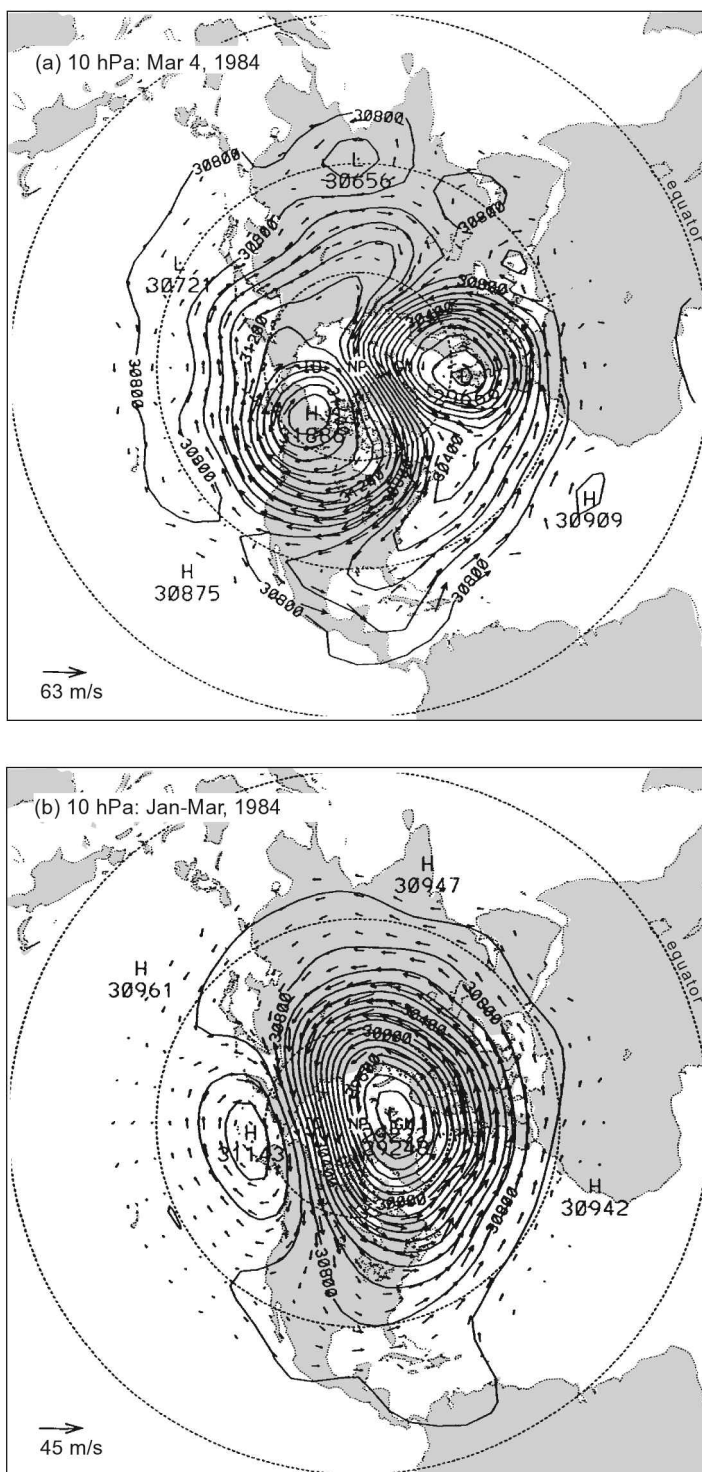


Figure 1.10 As in Fig. 1.9, but for the 10-hPa isobaric surface.

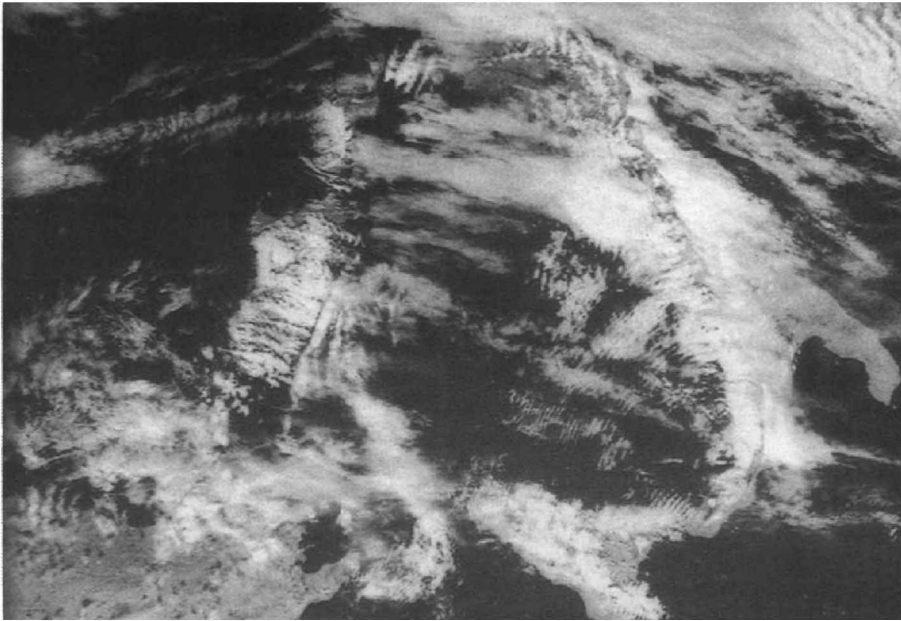


Figure 1.11 Satellite image of wavy cloud patterns found downwind of mountainous terrain. From Scorer (1986). Reproduced with permission of Ellis Horwood Ltd.

large excursions in latitude. Air is then transported from one radiative environment to another (e.g., from the sunlit tropics into polar darkness).

In addition to large-scale features, like those described in the preceding paragraphs, the circulation is also disturbed on smaller dimensions that are not resolved in the global analyses shown in Figs. 1.9 and 1.10. Known as *gravity waves*, these small-scale disturbances owe their existence to buoyancy and the stratification of mass. Gravity waves are manifested in wavy patterns that appear in layered clouds. They are often observed from the ground and in satellite imagery like that shown in Fig. 1.11. Like planetary waves, gravity waves contain both transient and steady components. Consequently, they are present even in time-mean fields.

1.2.4 Trace constituents

Beyond its primary constituents, air contains a variety of trace species. Although present in minor abundances, several play key roles in radiative and chemical processes. Perhaps the simplest is CO_2 because it is chemically inert away from the Earth's surface and therefore well-mixed throughout the homosphere. Like N_2 and O_2 , carbon dioxide has a nearly uniform mixing ratio, $r_{\text{CO}_2} \cong 380$ ppmv in 2010. However, unlike the primary constituents of air, CO_2 is coupled to human activities.

Carbon Dioxide

Involved in chemical and biological processes, CO_2 is produced and destroyed naturally near the Earth's surface. It is produced through oxidation processes, like cell respiration and soil respiration. The latter is associated with the bacterial

decomposition of organic matter under aerobic conditions (aerobic fermentation). These processes absorb oxygen and convert it into carbon dioxide. CO₂ is also produced under anaerobic conditions, at deeper soil levels. Involving similar pathways, anaerobic fermentation likewise involves the bacterial decomposition of organic matter, whereby carbohydrates are oxidized to produce CO₂ and hydrocarbons like methane. CO₂ is destroyed through reverse pathways, reduction processes – notably photosynthesis. Those processes absorb carbon dioxide and convert it into carbohydrate and back into oxygen. Together, these natural processes lead to emission of ~150 gigatons of carbon per year (150 GtC/yr), offset by about as much absorption (see Fig. 17.11). Almost half of the O₂ emitted to and CO₂ absorbed from the atmosphere are estimated to come from *phytoplankton*, marine biota that form the underpinnings of the food chain and cycle CO₂ into O₂.

The relationship of CO₂ to temperature is documented in prehistorical records of the Earth's climate. Glacial ice cores, drilled from great depths, provide a record of atmospheric composition dating far into the past. In concert with isotopic information on atmospheric temperature, that record suggests a link between atmospheric CO₂ and global temperature.

During previous climates of the Earth, these quantities varied in a systematically related fashion. Figure 1.12a compares records of CO₂ and temperature, inferred from an Antarctic ice core that extends back 400,000 years. Minima in CO₂ (dashed) were achieved when its mixing ratio had decreased below 200 ppmv. Those features coincide with anomalously-cold conditions, when, simultaneously, temperature (solid) had decreased by some 8 K. Both coincide with *glacial periods*, when ice volume (Fig. 1.12b) maximized some 20,000, 140,000, 250,000, and 350,000 years ago. The periods of ice maxima were followed by periods of warming, *interglacial periods*, when CO₂ and temperature increased sharply. During interglacials, there was a sharp recession of ice volume. Compared with the gradual approach to glacial maxima, changes of CO₂, temperature, and ice during the swings to glacial minima are abrupt. Close inspection of the records in Fig. 1.12a also indicates a small but repeated separation between contemporaneous features, with CO₂ lagging temperature by 500–1000 years. The implied relationship of CO₂ to temperature is manifest even on much shorter time scales (Sec. 1.6.2).

The veracity of proxy records like those in Fig. 1.12 is clouded by uncertainties (ibid). Notable is the long-term stability of gases that are trapped inside ice, along with their diffusion between layers. Those uncertainties limit temporal resolution, which hampers the discrimination to individual periods. They also act to homogenize properties, limiting excursions in older portions of the record.⁵ Equally important is the issue of causality. While illustrating the interdependence of CO₂ and temperature, these proxies provide little insight into which property produced changes in the other. Perhaps relevant is that they also leave undocumented contemporaneous changes in the most important radiatively-active species, water vapor and cloud, which are far more influential (Chap. 8).

More recent records evidence a human contribution to the budget of CO₂. Since the dawn of the industrial era (late eighteenth-century), the combustion of fossil fuel has steadily increased the rate at which carbon dioxide is introduced into the atmosphere. Augmenting that source is biomass destruction, notably, in connection

⁵ The signature of diffusion is apparent in Fig. 1.12, wherein older sections of the record become increasingly blurred.

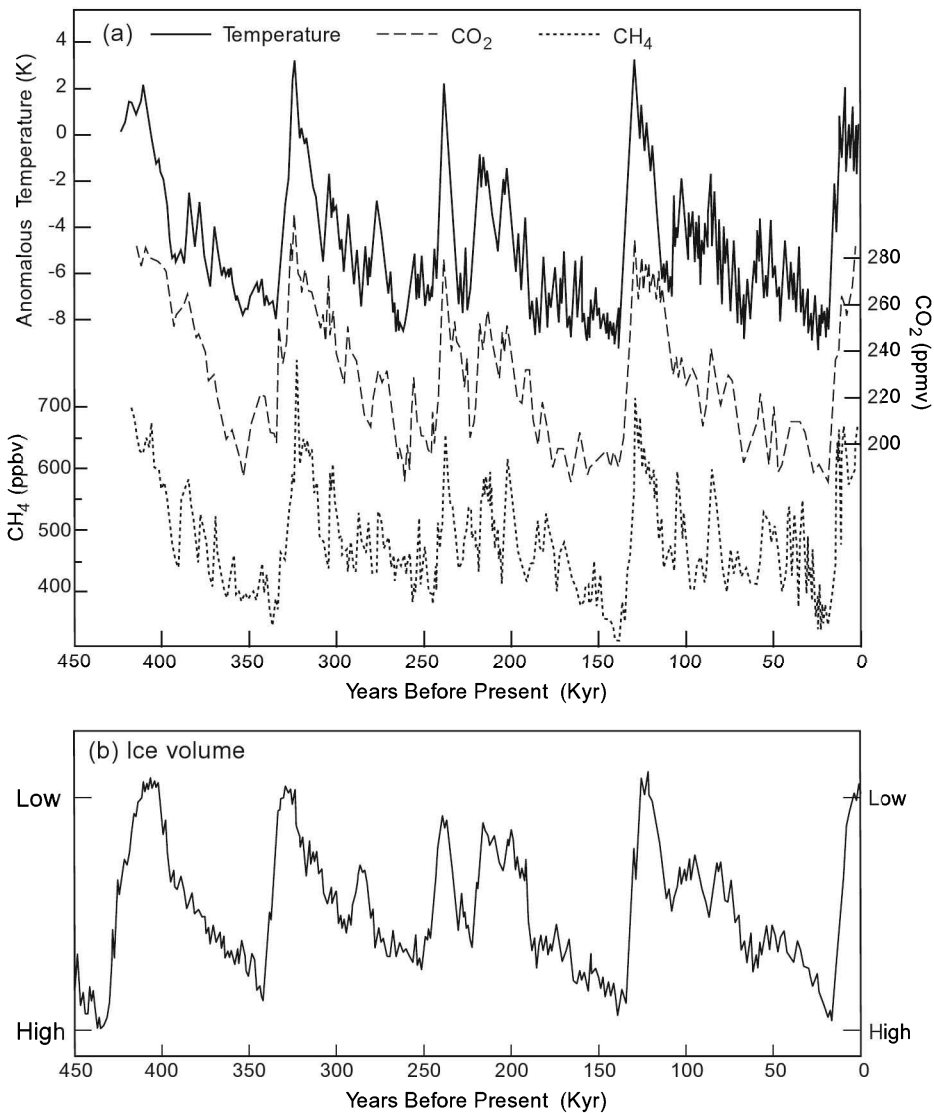


Figure 1.12 (a) Reconstructions of atmospheric temperature (solid), carbon dioxide (dashed), and methane (dotted) as functions of time before present, inferred from ice core drilled at Vostok Antarctica. (b) Reconstruction of global ice volume, inferred from isotopic analysis of ^{18}O in sedimentary foraminifera. *Sources:* Petit et al. (1999); Lisiecki and Raymo (2005); <http://www.globalchange.gov>.

with the clearing of dense tropical rainforest for timber and agriculture. (This process produces CO_2 either directly, through burning of vegetation, or indirectly, through its subsequent decomposition.) Interactions with the ocean and the biosphere make the budget of CO_2 complex. Nevertheless, the involvement of human activities is strongly suggested by observed changes.

Plotted in Fig. 1.13 is the record of CO_2 mixing ratio over the last 1000 years, inferred from ice cores over Antarctica (symbols). According to this proxy, r_{CO_2} varied

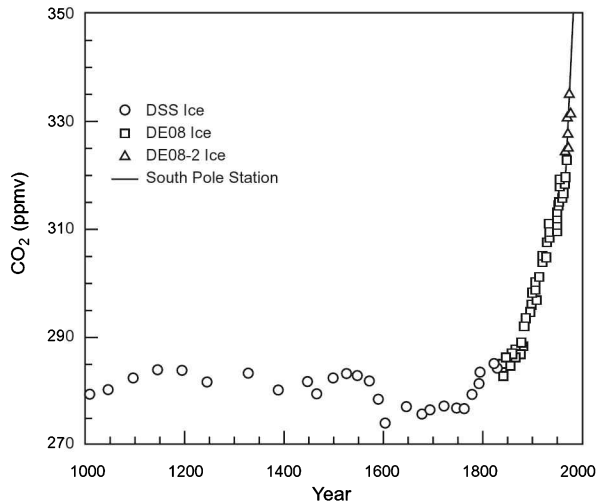


Figure 1.13 Mixing ratio of proxy CO_2 (ppmv) during the preceding millennium, inferred from Antarctic ice cores (symbols). Superimposed is the more recent instrumental record of atmospheric CO_2 (solid). After Ethridge et al. (1996).

during most of the last millennium by less than 10 ppmv. It hovered about a nearly constant value of 280 ppmv. After ~ 1800 , however, r_{CO_2} began a steep and monotonic increase, one that has prevailed to date. The proxy evidence is consistent with nearby instrumental measurements of CO_2 , which became available in the twentieth century (solid). Jointly, these records describe a modern increase that has brought r_{CO_2} to values in excess of 380 ppmv, about 35% higher than pre-industrial values in the proxy record.

The ice record of CO_2 in Fig. 1.12a is expanded in Fig. 1.14, which plots, at higher resolution, r_{CO_2} over the last 150 years (squares). Data have been smoothed to exclude variability with periods shorter than two decades. The monotonic increase dates back to around 1850, with interruptions during the 1880s and 1890s and, more conspicuously, during the 1940s and 1950s. If anything, r_{CO_2} during those intervals decreased. The reconstructed evolution is, like the longer proxy record (Fig. 1.13), consistent with more recent instrumental measurements of CO_2 (triangles).

The upward trend of CO_2 is commonly ascribed to emission by human activities. Support for this interpretation comes from isotopes of carbon. Carbon 13, like carbon 12, is stable. It represents about 1% of the isotopic composition of CO_2 . However, its concentration varies between reservoirs of carbon. Vegetation and ancestral carbon, fossil fuel, are slightly leaner in ^{13}C than is the atmosphere.⁶ Also plotted in Fig. 1.14 is the relative concentration of atmospheric ^{13}C ,

$$\delta^{13}\text{C} = \frac{(^{13}\text{C}/^{12}\text{C})}{(^{13}\text{C}/^{12}\text{C})_{\text{ref}}} - 1$$

in parts per thousand (‰), referenced against a standard value (solid circles). Proxy evidence of $\delta^{13}\text{C}$ is more variable than that of r_{CO_2} . It also has little overlap with the

⁶ Reflecting increased efficiency of photosynthesis with the lighter form of carbon.

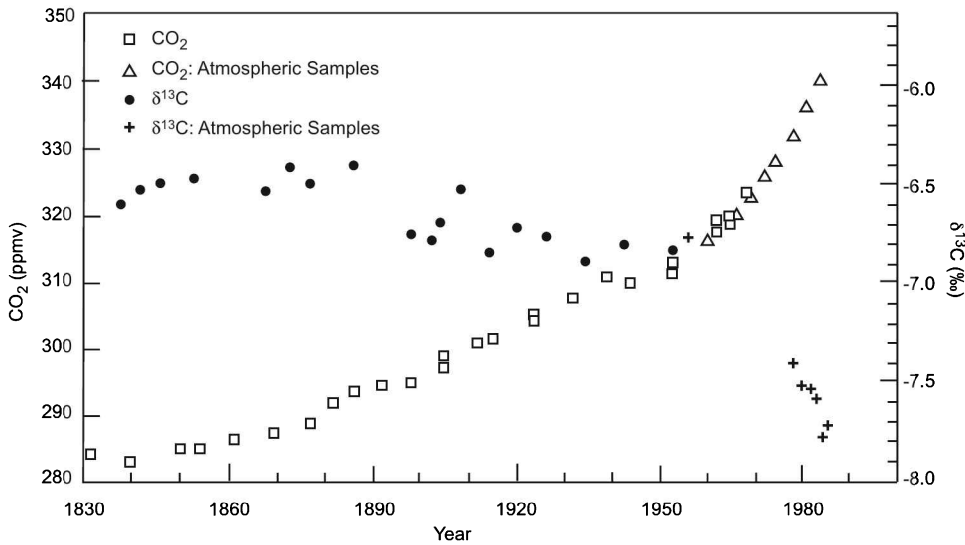


Figure 1.14 Mixing ratio of proxy CO_2 (ppmv) during the last two centuries, inferred from Antarctic ice cores (squares). Superimposed is the contemporaneous instrumental record of r_{CO_2} over the South Pole (triangles). Also plotted is the relative concentration of isotopic carbon (per thousand) $\delta^{13}\text{C} = \frac{(^{13}\text{C}/^{12}\text{C})}{(^{13}\text{C}/^{12}\text{C})_{ref}} - 1$ (solid circles), along with recent instrumental observations (crosses). *Sources:* Etheridge et al. (1996); Friedli et al. (1986).

more recent instrumental record (crosses). Nevertheless, reconstructed $\delta^{13}\text{C}$ decreased over the last two centuries, mirroring the contemporaneous increase of r_{CO_2} .

The decrease of $\delta^{13}\text{C}$, together with the increase of r_{CO_2} , reflects the addition of CO_2 that is ^{13}C lean. This feature is consistent with the combustion of fossil fuel, as well as biomass destruction. It is equally consistent, however, with the decomposition of organic matter derived from vegetation. Thus, associating the decrease of $\delta^{13}\text{C}$ to the combustion of fossil fuel requires the exclusion of other sources that are ^{13}C lean. In particular, it relies on CO_2 emission from the ocean, which overshadows other sources of CO_2 (Sec. 17.3), having the same isotopic composition as the atmosphere (which would then be left unchanged). Only then can the decrease of $\delta^{13}\text{C}$ be isolated to continental sources, which are weaker and, in particular, to the combustion of fossil fuel, which is an order of magnitude weaker. Yet, the isotopic composition of marine organic matter is influenced by a variety of biological and environmental factors (Francois et al., 1993; Goericke et al., 1994). Through those factors, $\delta^{13}\text{C}$ in the upper ocean varies significantly. Along with transport from the deep ocean, which is likewise uncertain, they leave the magnitude and composition of ocean emission poorly understood (Sec. 1.6.2).

The instrumental record is expanded in Fig. 1.15, which plots r_{CO_2} observed at Mauna Loa during the latter half of the twentieth century. Situated in the central Pacific and several kilometers above sea level, Mauna Loa measurements are removed from continental and urban sources. Accordingly, they reflect CO_2 that has become well mixed across the troposphere. An annual variation of about ± 3 ppmv is synchronized

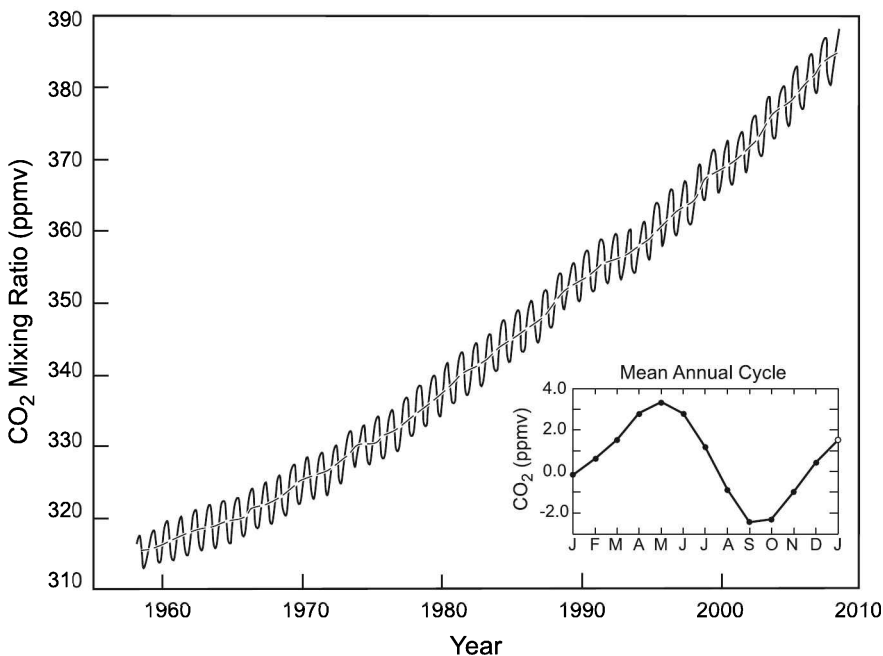


Figure 1.15 Mixing ratio of carbon dioxide over Mauna Loa, Hawaii. Superimposed is the mean annual variation. *Source:* www.esrl.noaa.gov/gmd.

with the growing season of the Northern Hemisphere: CO₂ decreases during May–September, reflecting absorption by vegetation. The annual variation, however, is not cyclic. More CO₂ is emitted during part of the year, when sources prevail over sinks and r_{CO_2} increases, than is absorbed during the other part of the year, when sinks prevail over sources and r_{CO_2} decreases. This yields a hysteresis in the annual variation: CO₂ in one January is ~ 1.5 ppmv greater than CO₂ in the preceding January. During successive years, the hysteresis accumulates to form a steady upward trend, the same trend apparent over Antarctica. Through that trend, CO₂ has increased by more than 20% in just half a century.

The rapid increase of CO₂ during the industrial era has raised concern over global warming because of the role carbon dioxide plays in trapping radiant energy near the Earth's surface. Developed in Chap. 8, this role makes carbon dioxide a *greenhouse gas*. The implication to temperature of increasing CO₂ was first noted by the Swedish chemist, Arrhenius (1894). For Arrhenius, global warming was, ironically, but a curiosity. His real interest was to explain the ice ages, behavior he pursued through a decrease of CO₂. In a crude but challenging calculation of radiative energy transfer, Arrhenius estimated that halving CO₂ would lead to cooling of 4–5 K, sufficient to form an ice age. Through the same calculation, he estimated that a doubling of CO₂ would lead to warming of 5–6 K. For the rate at which CO₂ is currently increasing, $\sim 20\%$ during the last 5 decades, this translates into a warming rate of about 0.25 K/decade.

Arrhenius was unconcerned by the latter possibility because, at the rate of emission then, he estimated that such levels of CO₂ would not be achieved for millennia. The possibility became more tangible in his subsequent book (Arrhenius, 1908), wherein

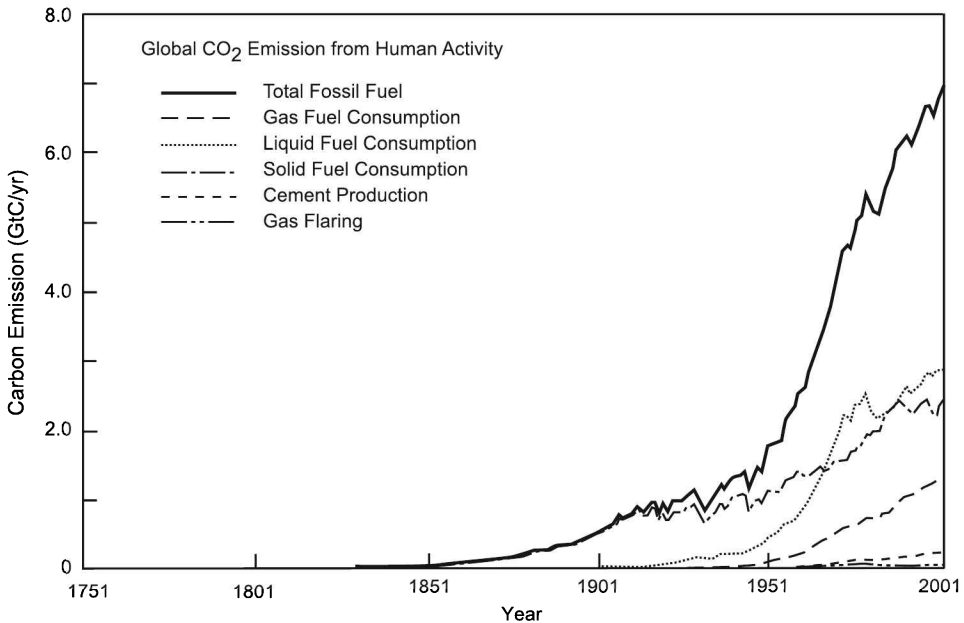


Figure 1.16 Anthropogenic emission of CO₂, in gigatons of carbon per year (GtC/yr), as total and from major sources. *Source:* Marland et al. (2008).

CO₂ was foreseen to double in only centuries. Other researchers doubted that it would ever occur. This position followed from radiative considerations: (1) Absorption by CO₂ was thought to involve the same wavelengths of radiation as water vapor, which is far more abundant and fully absorbs radiation at those wavelengths. (2) The atmosphere already contained sufficient CO₂ to absorb all of the radiation possible; increased CO₂ would therefore be inconsequential. Even exclusive of radiative considerations, most of the CO₂ emission at the turn of the nineteenth century was thought to be absorbed by the ocean. What had not been evaluated was the capacity of the ocean to keep pace with an increasing rate of CO₂ emission.⁷

Plotted in Fig. 1.16 is anthropogenic emission of CO₂ in gigatons of carbon per year (GtC/yr). Around 1850, CO₂ emission ramped up exponentially, approaching 1 GtC/yr at the time of Arrhenius. Following World War I and during the Great Depression, it increased slower. Then, after World War II, CO₂ emission increased sharply, chiefly through combustion of liquid fuel that supports transportation. Almost as great is emission from the combustion of solid fuel in power generation and industry. It is noteworthy that, during the 1940s and 1950s, when emission from fossil fuel accelerated, proxy CO₂ inferred from the ice-core record (Fig. 1.14) did not. If anything, r_{CO_2} during that period decreased. Current rates of anthropogenic emission exceed 7 GtC₂/yr. Although increasing, this is still only about 4% of total CO₂ emission, which is dominated by natural sources; cf. Fig. 17.11. Consequently, even a minor imbalance

⁷ The growth of anthropogenic emission parallels the expansion of humans, who share basic needs of food, heat, and transportation. Those needs all require energy. At the time of Arrhenius, the human population was 1.5 billion. Today, it exceeds 6 billion. By the second half of the twenty-first century, it is projected to exceed 10 billion.

between natural sources and sinks can overshadow the anthropogenic component of CO₂ emission.

The concern over increasing CO₂ is supported, in part, by large-scale numerical simulations. Global Climate Models (GCMs) are used to study climate by including a wide array of physical processes. Many can be represented only crudely, with more than a few represented through ad hoc treatment. Although central to many of the considerations, those processes must be treated through crude parameterization because either (1) they cannot be resolved in the computation, (2) they are not adequately observed, or (3) the governing equations that describe them are not even known. Owing to such uncertainties, GCMs differ widely in detail. Nevertheless, on basic features, they are in general agreement. Current models predict that a doubling of CO₂ would lead to warming of global-mean temperature by 2–6 K (IPCC, 2007). Those projections are broadly consistent with Arrhenius' simple calculation, which considered only fundamental issues surrounding the transfer of radiative energy.

Although differing between models, projected changes of temperature are sufficiently great to imply important changes to the Earth's climate (e.g., by melting of continental ice and increasing sea level). In the stratosphere, where CO₂ dominates infrared (IR) cooling to space, temperature decreases as large as 10 K have been suggested. They, in turn, could alter other radiatively active constituents such as ozone. Distinguishing the anthropogenic component of such changes from the natural component will be essential to correctly interpret observed changes and to understand how they are likely to evolve. This is especially true for changes of regional climate, for which model projections are far less reliable than for global-mean temperature (Sec. 8.7.3).

Water vapor

The uniform composition of dry air in the homosphere is not to be confused with the distributions of trace species such as water vapor and ozone. These constituents are highly variable because they are not merely redistributed by atmospheric motion, which would eventually homogenize them. Instead, water vapor and ozone are produced in some regions and destroyed in others. By transporting them from their source regions to their sink regions, the circulation exerts an important influence on these species and makes their distributions dynamic.

Owing to its involvement in radiative processes, cloud formation, and in exchanges of energy with the ocean, water vapor is the single most important trace species in the atmosphere. Its strong absorptivity in the IR, developed in Chap. 8, makes it also the single most important greenhouse gas. The zonal-mean distribution of water vapor is shown in Fig. 1.17 as a function of latitude and altitude. Water vapor is confined almost exclusively to the troposphere. Its zonal-mean mixing ratio \bar{r}_{H_2O} decreases steadily with altitude, from a maximum of about 20 g kg⁻¹ (0.020) at the surface in the tropics to a minimum of less than 1 ppm at the tropopause. The absolute concentration of water vapor, or *absolute humidity* $\bar{\rho}_{H_2O}$ (shaded), decreases with altitude even more rapidly. From (1.14), the density of the *i*th constituent is just its mixing ratio times the density of dry air

$$\rho_i = r_i \rho_d. \quad (1.24)$$

Because ρ_d decreases exponentially with altitude, water vapor tends to be concentrated in the lowest 2 km of the atmosphere. The zonal-mean mixing ratio also decreases with

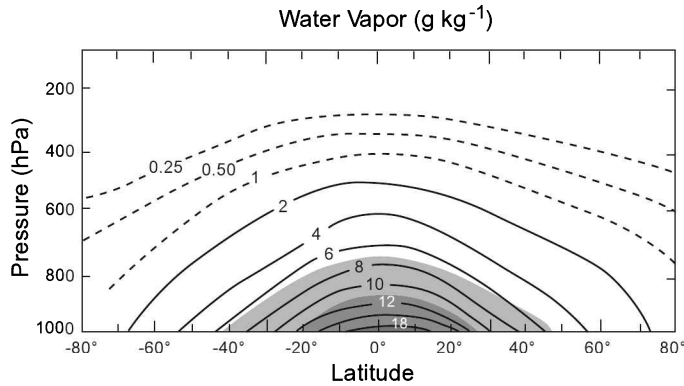


Figure 1.17 Zonal-mean mixing ratio of water vapor (contoured) and density of water vapor or absolute humidity (shaded), as functions of latitude and pressure. Shading levels correspond to 20%, 40%, and 60% of the maximum value. *Source:* Oort and Peixoto (1983).

latitude, falling to under 5 g kg^{-1} poleward of 60° . These characteristics of water vapor reflect its production over warm tropical ocean at the Earth's surface, redistribution by the atmospheric circulation, and destruction at altitude and at middle and high latitudes through condensation and precipitation.

Owing to those production and destruction mechanisms and the rapid transport of air between source and sink regions, tropospheric water vapor is short-lived. A characteristic lifetime, the time for $r_{\text{H}_2\text{O}}$ inside an individual air parcel to change significantly, is of order days. Every few days, an air parcel encounters a warm ocean surface, where it absorbs moisture through evaporation, or a region of cloud, where it loses water vapor through condensation and precipitation.

Most of the water vapor in Fig. 1.17 originates near the equator from warm ocean surface. Consequently, transport by the circulation plays a key role in determining the mean distribution $\bar{r}_{\text{H}_2\text{O}}$. Vertical and horizontal air motion are referred to respectively as *convection* and *advection*. Each contributes to the redistribution of $r_{\text{H}_2\text{O}}$. Introduced at the surface of the tropical atmosphere, water vapor is carried upward by deep convective cells. It is carried horizontally by large-scale eddies, which disperse $r_{\text{H}_2\text{O}}$ across the globe. Some bodies of air escape production and destruction long enough for $r_{\text{H}_2\text{O}}$ to be conserved. Mixing ratio is then rearranged as a tracer.

Figure 1.18 presents, for the day shown in Fig. 1.9, an image from the $6.3\text{-}\mu\text{m}$ water vapor channel of the geostationary satellite Meteosat-2, which observed the Earth from above the Greenwich meridian (see Fig. 1.29 for geographical landmarks). The gray scale in Fig. 1.18 represents cold emission temperature (high altitude) as bright and warm emission temperature (low altitude) as dark. The water vapor column is optically thick at this wavelength. Outgoing radiation therefore derives from H_2O emission only at the highest levels of the water vapor column. For this reason, behavior in Fig. 1.18 characterizes the top of the moisture layer. Bright regions indicate moisture at high altitude, corresponding to deep convective displacements of surface air. Conversely, dark regions indicate moisture that remains close to the Earth's surface. Together, those structures reflect the horizontal distribution of $r_{\text{H}_2\text{O}}$.

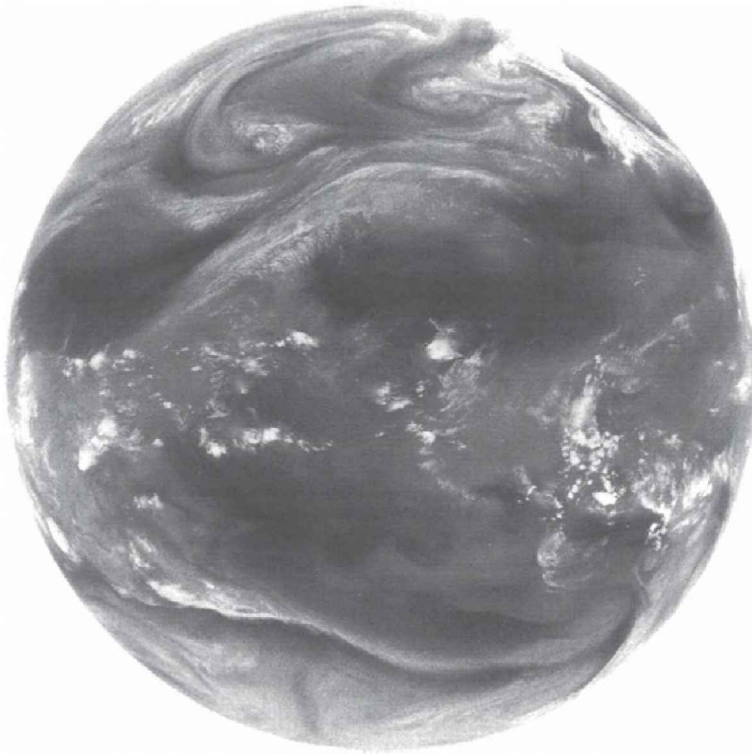


Figure 1.18 Water vapor image on March 4, 1984, from the $6.3 \mu\text{m}$ channel of Meteosat-2, which is in geostationary orbit over the Greenwich meridian. Gray scale displays equivalent blackbody temperature from warmest (black) to coldest (white). (See Fig. 1.29 for geographical landmarks). Supplied by the European Space Agency.

Unlike its mean distribution in Fig. 1.17, which is fairly smooth, the global distribution of water vapor on an individual day is quite variable. Especially dynamic is the distribution in the tropics, where the moisture pattern is granular. There, water vapor has been displaced upward by deep convective cells that have horizontal dimensions as short as tens of kilometers and time scales of only hours. At middle and high latitudes, the pattern is smoother, but still complex. Swirls of light and dark mark bodies of air that are rich and lean in water vapor, respectively. They reflect air which originated in tropical and extratropical regions but which has been rearranged by the circulation. The local abundance reflects the history of the air parcel residing at that location, namely where that parcel has been and what processes influencing water vapor have acted on it. A plume of high humidity stretches northeastward from deep convection over the Amazon basin (cf. Fig. 1.29). It extends across the Atlantic and into Africa, where it joins a plume of drier air that is being drawn southward behind a cyclone in the eastern Atlantic (cf. Fig. 1.9a). In the Southern Hemisphere, a band of high humidity is sharply delineated from neighboring dry air along a front that spans the south Atlantic.

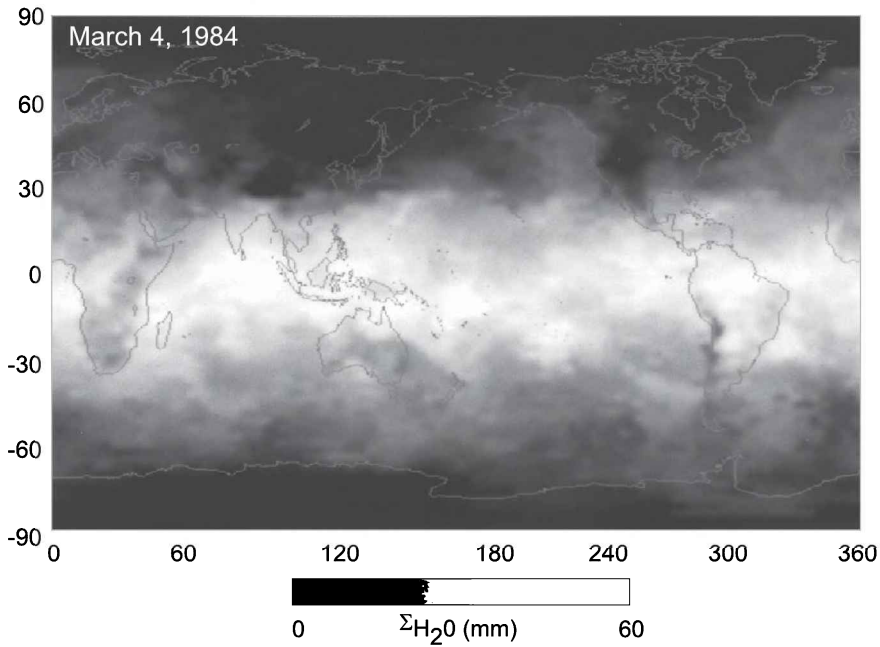


Figure 1.19 Global distribution of the column abundance of water vapor, or *total precipitable water vapor*, on March 4, 1984, derived from the TIROS Operational Vertical Sounder (TOVS). Data courtesy of I. Wittmeyer and T. Vonderharr.

More relevant to radiative processes than the local concentration is the total abundance of a species over a particular horizontal position. The *column abundance*

$$\begin{aligned}\Sigma_i &= \int_0^{\infty} \rho_i dz \\ &= \frac{1}{g} \int_0^{p_s} r_i dp,\end{aligned}\tag{1.25}$$

describes the mass of the i th species contained by an atmospheric column of unit cross-sectional area. Figure 1.19 displays the distribution of Σ_{H_2O} , referred to as *total precipitable water vapor*. Expressed in millimeters of liquid water, it is presented for the same day as in Fig. 1.18. Σ_{H_2O} is sharply confined to the tropics, resembling the distribution of temperature. It is distributed more uniformly than r_{H_2O} (Fig. 1.18). In part, this feature of precipitable water vapor follows from the concentration of ρ_{H_2O} in the lowest 2 km. But it also indicates that, at higher altitude, even deep convective towers that punctuate Fig. 1.18 contain comparatively little water in vapor phase.

Together, the zonal-mean and horizontal distributions illustrate the prevailing mechanisms that control atmospheric water vapor. Large values of \bar{r}_{H_2O} at the Earth's surface in the tropics reflect production of water vapor through evaporation of warm tropical ocean. Only inside convective towers and in analogous features of greater dimension are large mixing ratios found far above the ground. Even there, vertical transport of H_2O is limited by thermodynamic constraints. Developed in Chap. 5, they

prevent water vapor from reaching great altitude, where it would suffer photodissociation by UV and eventually be lost to space through escape of atomic hydrogen (Sec. 1.2.2).

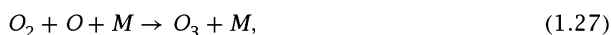
Ozone

O₃ is another radiatively active trace gas. It plays an essential role in supporting life at the Earth's surface. By intercepting harmful UV radiation, ozone allows life as we know it to exist.⁸ The evolution of the Earth's atmosphere and the formation of the ozone layer are, in fact, thought to be closely related to the development of life on Earth.

Geological evidence suggests that primitive forms of plant life developed *in aqua*, deep in the ocean, at a time when the Earth's atmosphere contained little or no oxygen. Damaging UV radiation then passed freely to the Earth's surface. Through photosynthesis, these early forms of life are thought to have liberated oxygen, which then passed to the surface. There, it was photo-dissociated by UV radiation in the reaction



Atomic oxygen produced by (1.26) could then recombine with O₂ to form ozone in the thermomolecular reaction



where *M* represents a third body needed to carry off excess energy liberated by the combination of O and O₂. Ozone created in (1.27) is dissociated by UV radiation according to the reaction



If third bodies are abundant, atomic oxygen produced by (1.28) recombines almost immediately with O₂ in (1.27) to re-form ozone. Thus reactions (1.27) and (1.28) constitute a "closed cycle," one that involves no net loss of components. Because the only result is the absorption of solar energy, this cycle can process UV radiation efficiently. By removing harmful UV from the solar spectrum, ozone is thought to have allowed life to advance upward to the ocean surface. There, it had greater access to visible radiation, could produce more oxygen through photosynthesis, and, consequently, was able to evolve into more sophisticated forms.

Ozone also absorbs IR radiation, a property that makes it a greenhouse gas. Its contribution to increased warming of the Earth's surface is thought to be about 20% of the contribution from CO₂ (Fig. 8.30). The increased warming from O₃ follows from tropospheric ozone, which, like CO₂, is increasing. It represents as much as a third of the total ozone column (Fig. 18.2).

The zonal-mean distribution of ozone mixing ratio is plotted in Fig. 1.20 as a function of latitude and altitude. Whereas atmospheric water vapor is confined chiefly to the troposphere, ozone is concentrated in the stratosphere. Ozone mixing ratio increases sharply above the tropopause, reaching a maximum of about 10 ppmv near

⁸ UV exposure is strongly correlated to common forms of skin cancer, basal cell and squamous cell carcinoma, which have low rates of mortality. Its correlation to melanoma, which has a high rate of mortality, is weaker and remains under debate (Osterlind et al, 2006; Cascinellu and Marchesini, 2008; Bauer, 2010).

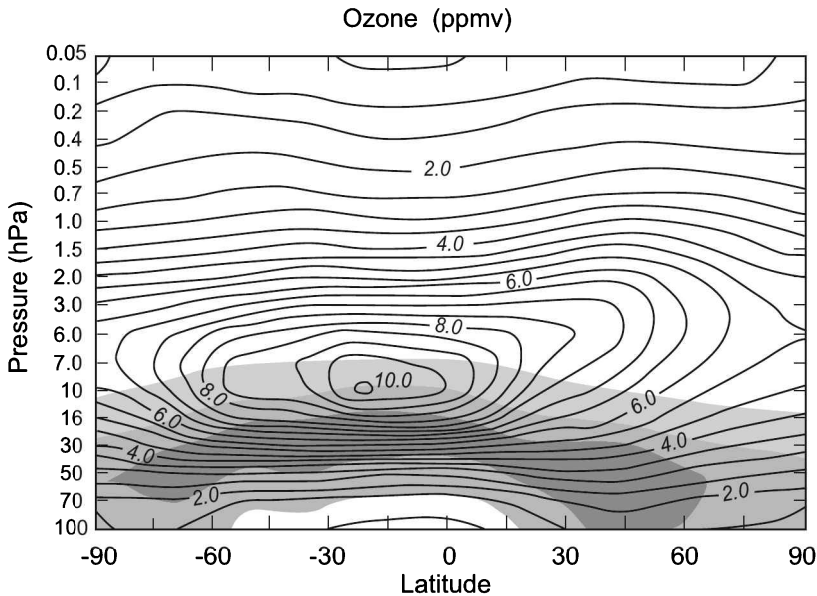


Figure 1.20 Zonal-mean mixing ratio of ozone (contoured) and density of ozone (shaded) averaged over January–February 1979, as functions of latitude and pressure, obtained from the Limb Infrared Monitor of the Stratosphere (LIMS) on board Nimbus-7. Shading levels correspond to 20%, 40%, and 60% of the maximum value.

30 km (10 hPa). The zonal-mean ozone mixing ratio \bar{r}_{O_3} is largest in the tropics, where the flux of UV and photodissociation of O_2 are large.

The photochemical lifetime of ozone varies sharply with altitude. In the lower stratosphere, ozone has a photochemical lifetime of several weeks. Because this is long compared with the characteristic time scale of air motion (~ 1 day), r_{O_3} behaves as a tracer at these altitudes. For the same reason, its distribution there is controlled by dynamical influences. The lifetime of ozone decreases upward, shortening to of order 1 day by 30 km and only an hour by the stratopause. For this reason, the distribution of ozone in the upper stratosphere and mesosphere is controlled by photochemical influences.

The troposphere serves as a sink of stratospheric ozone. Should ozone find its way into the troposphere, it is quickly destroyed. Its water-solubility enables O_3 to be absorbed partially by convective systems. Through vertical transport, they transfer ozone to the Earth's surface, where it is consumed in a variety of oxidation processes.

Even though its mixing ratio \bar{r}_{O_3} maximizes near 30 km, atmospheric ozone is concentrated in the lower stratosphere (Fig. 1.20). Because air density decreases exponentially with altitude, the density of ozone $\bar{\rho}_{O_3}$ (shaded) is concentrated at altitudes of 10–20 km (1.24), not far above the tropopause. The largest values are found in a shallow layer near 30 hPa in the tropics, which descends and deepens in extratropical regions. The column abundance, or *total ozone* Σ_{O_3} , is expressed in Dobson units (DU). One DU measures, in thousandths of a centimeter, the depth the ozone column

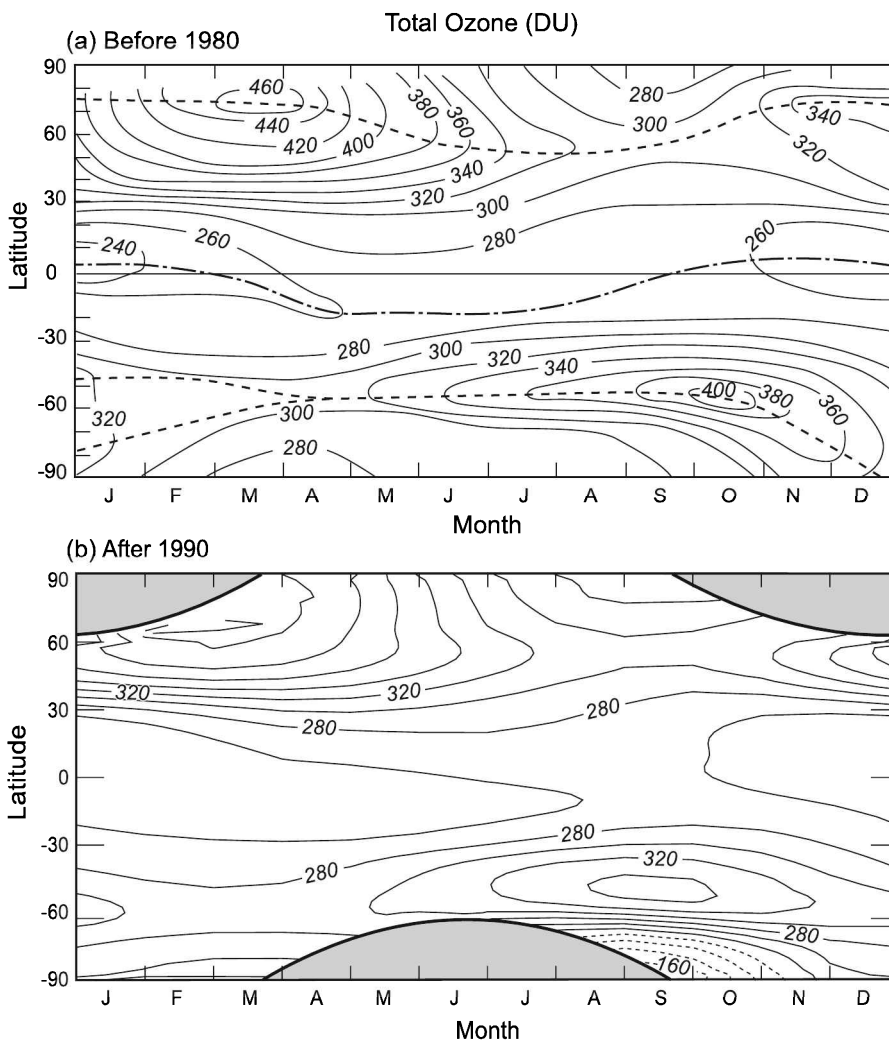


Figure 1.21 Zonal-mean column abundance of ozone, or *total ozone*, as a function of latitude and month. (a) Based on the historical record prior to 1980. From London (1980). (b) Based on satellite observations during the 1990s, after the formation of the Antarctic ozone hole, from the Total Ozone Mapping Spectrometer (TOMS). Values less than 220 DU (dotted) characterize the ozone hole. Observations not available in regions of darkness (shaded).

would assume if brought to standard temperature and pressure. Figure 1.21a plots the zonally averaged column abundance of ozone, as a function of latitude and season, based on observations before 1980. Values of $\bar{\Sigma}_{O_3}$ range from about 250 DU near the equator to in excess of 400 DU at high latitude. At standard temperature and pressure, the entire ozone column measures less than one half of one centimeter! Even though most stratospheric ozone is produced in the tropics, the greatest column abundance is found at middle and high latitudes. This peculiarity of the ozone distribution illustrates the importance of dynamics in shaping observed composition and structure.

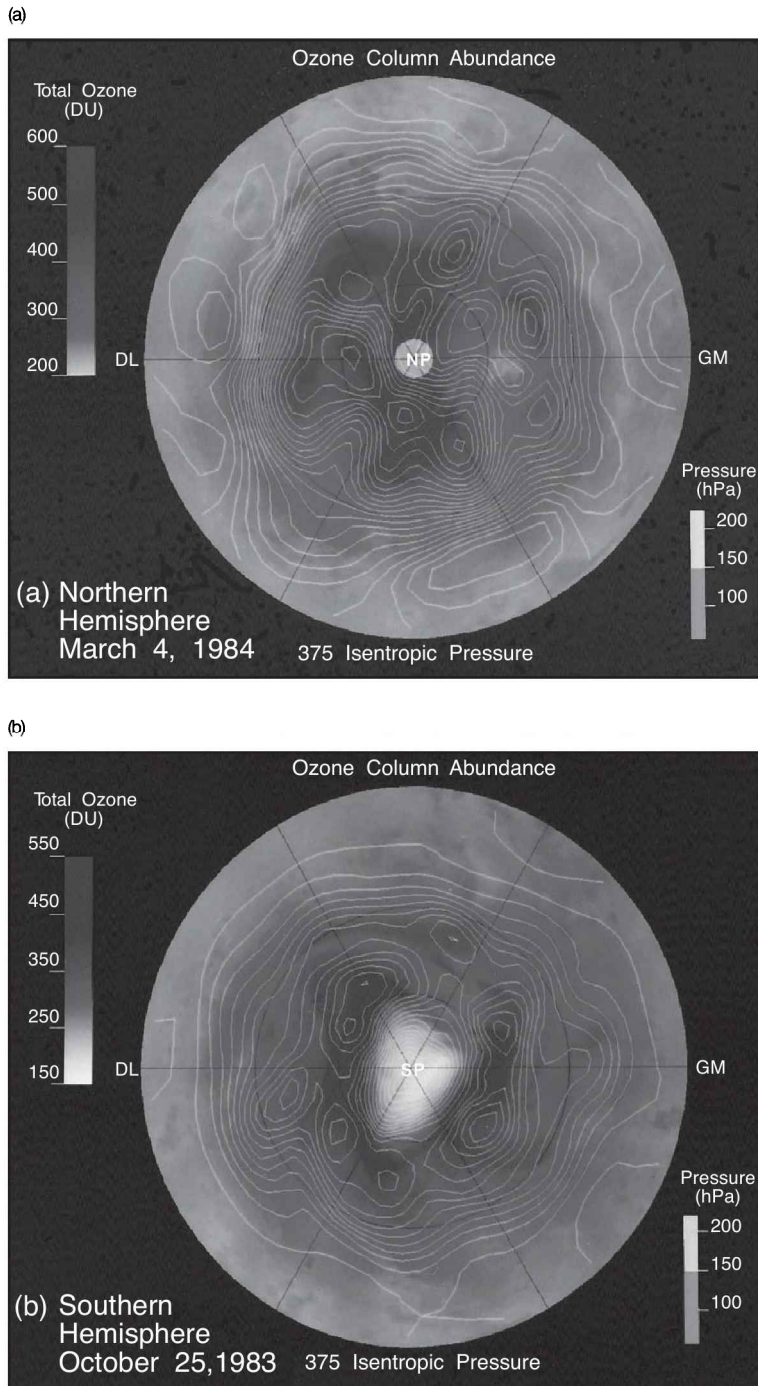


Figure 1.22 Distribution of total ozone (color) from the Total Ozone Mapping Spectrometer (TOMS) on board Nimbus-7 and pressure (contours) on the 375°K isentropic surface (see Sec. 2.4.1) over (a) the Northern Hemisphere on March 4, 1984 and (b) the Southern Hemisphere on October 25, 1983. See color plate section: Plate 1.

Compared in Fig. 1.21b is the same information, but from satellite observations during the 1990s (available only in sunlit regions). Distinguishing the behavior is anomalously low $\overline{\Sigma}_{O_3}$ over the Antarctic during Austral spring. Values less than 220 DU (dotted) then characterize the Antarctic ozone hole, which appeared around 1980. Reflecting strong depletion, they take a bite out of the annual cycle of Antarctic ozone, which, before 1980, achieved values nearly twice as great.

As is true for water vapor, the circulation plays a key role in determining the mean distribution of O_3 . It makes the global distribution on individual days dynamic. Figure 1.22 presents the distribution of Σ_{O_3} over the Northern and Southern Hemispheres on individual days, as observed by the Total Ozone Mapping Spectrometer (TOMS). The distribution over the Northern Hemisphere (Fig. 1.22a) has been arranged by the circulation into several anomalies in which Σ_{O_3} varies by as much as 100%. Figure 1.22a, which is contemporaneous with the 500-hPa circulation in Fig. 1.9a, reveals a strong correspondence between variations of Σ_{O_3} and synoptic weather systems in the troposphere. Both evolve on a time scale of a day. The distribution over the Southern Hemisphere (Fig. 1.22b) is distinguished by column abundance less than 200 DU (white), which delineates the “Antarctic ozone hole.” Anomalously low Σ_{O_3} develops over the South Pole each year during Austral spring. Forming inside the polar-night vortex, it then disappears some 2 months later. The formation of the ozone hole, which emerged in the 1980s, follows from increased levels of atmospheric halogens, like chlorine and bromine. The ozone hole’s disappearance each year occurs through dynamics.

Several other trace gases also figure importantly in radiative and chemical processes. These include species that are produced naturally, such as methane (CH_4), and species that are produced exclusively by human activities, such as chlorofluorocarbons (CFCs).

Methane

CH_4 forms through bacterial and surface processes that occur naturally. It is produced by digestion in grazing animals, by insects, and through anaerobic fermentation, the decomposition of organic matter under anoxic conditions (chiefly in wetlands). Methane is also produced by living plants, a process that accelerates with temperature (Keppler et al, 2006; Reid and Qaderi, 2009). Like natural processes that emit CO_2 , these natural sources of CH_4 are poorly documented. Methane is also produced by a variety of human activities such as mining, agriculture, landfills, industrial processes, and biomass destruction. Collectively, those anthropogenic sources may account for as much as half of CH_4 emission.

Methane is long-lived and therefore well mixed in the troposphere. There, it has a uniform mixing ratio of $r_{CH_4} \cong 1.7$ ppmv (Fig. 1.23). In the stratosphere, r_{CH_4} decreases with altitude as a result of oxidation. This process ultimately leads to the formation of stratospheric water vapor. It is thought to be responsible for the upward increase of r_{H_2O} in the stratosphere.

Methane also varied in prehistoric times. The record in ice cores is superimposed in Fig. 1.12 (dotted). Like carbon dioxide (dashed), the variation of methane tracks the variation of temperature (solid), at least on time scales longer than millennial. Consequently, CO_2 and CH_4 do not vary independently. Changes of one are attended by changes of the other. In the ice core record of prehistoric composition, the percentage change of CH_4 is 2.0–2.5 times the percentage change of CO_2 .

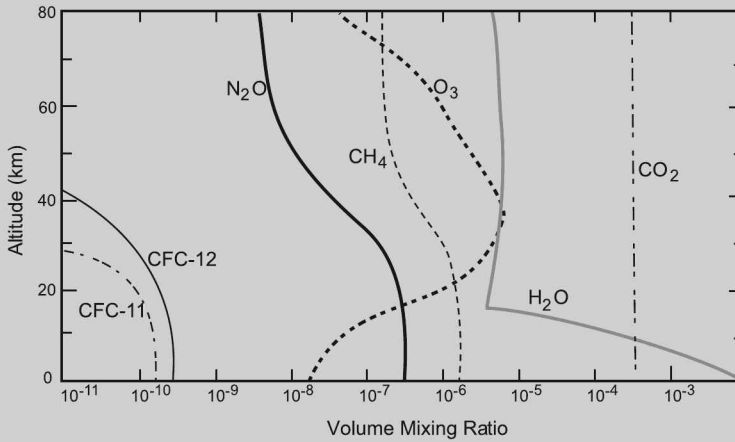


Figure 1.23 Mixing ratios of radiatively active trace species as functions of altitude. *Source:* Goody and Yung (1989).

Similar behavior characterizes changes during the modern era. As for carbon dioxide, methane concentrations are steadily increasing. The record inferred from ice cores over Antarctica and Greenland is presented in Fig. 1.24. It implies that, over much of the last 1000 yrs, r_{CH_4} hovered about a value of ~ 700 ppbv, when r_{CO_2} hovered about a value of ~ 280 ppmv (Fig. 1.13). However, like CO_2 , methane increased sharply after about 1850, now having values more than twice as great. In relation to anthropogenic emission, why CH_4 should increase contemporaneously with CO_2 is paradoxical. Methane has only a tenuous connection to the combustion of fossil fuel. Its production by agriculture has certainly increased. However, that increase has been in place for at least 7000 years, since the domestication of cattle. It is noteworthy that the decomposition of organic matter, which is intrinsic to the cycling

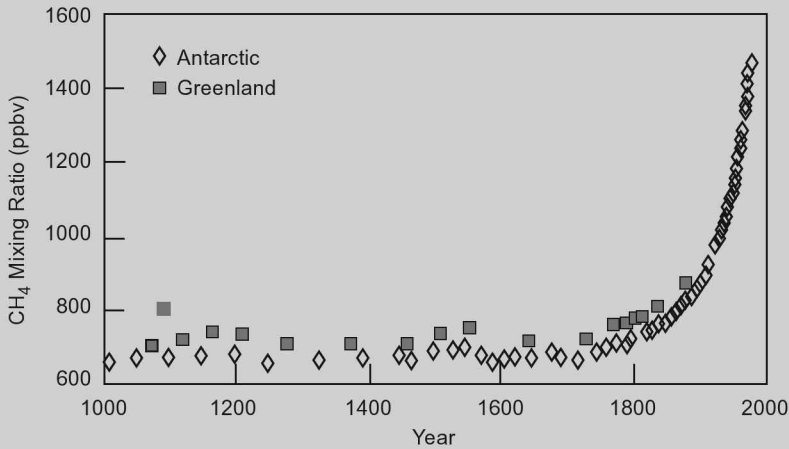


Figure 1.24 Mixing ratio of proxy CH_4 (ppbv) during the last millennium, inferred from ice cores from Antarctic and Greenland (symbols). *Source:* <http://cdiac.ornl.gov>, Etheridge et al. (CSIRO).

between oxygen and carbon dioxide, produces both methane and CO_2 . For many pathways, such as anaerobic fermentation, the concentration of emitted CH_4 is two to three times that of CO_2 .

Modern values of r_{CH_4} represent an increase over pre-industrial values of about 100%. This is three times the percentage increase of CO_2 during the same period. In the ice core record of prehistoric composition (Fig. 1.12), the relative change of CH_4 to CO_2 is 2.0–2.5. Although the ratio of modern changes exceeds that in the prehistoric record, the two are broadly consistent.

Like CO_2 , CH_4 is a greenhouse gas. It is less abundant than carbon dioxide, but more effective at trapping IR radiant energy. Per unit mass, methane is some 25 times more effective. Despite its much smaller abundance, this feature enables methane to contribute about a fourth as much increased warming of the Earth's surface as follows from elevated carbon dioxide (Fig. 8.30). Methane is also involved in the complex photochemistry of ozone (Sec. 18.2).

After the turn of the century, the rate at which methane is increasing slowed significantly. Estimates of anthropogenic emission did not (<http://www-naweb.iaea.org/naifa/aph/stories/2008-atmospheric-methane.html>; 14.11.10). In addition to ongoing production, large reserves of methane are trapped in frozen tundra beneath Arctic ice, notably over Siberia. It has been suggested that the release of those reserves, if Arctic ice recedes, would sharply magnify CH_4 emission. The attendant warming would, in turn, reinforce the melting of Arctic ice, representing a positive feedback. How this suggestion fits with the seasonality of methane emission over Arctic tundra is unclear. During the months of autumn, when tundra freezes, CH_4 emission is as great as during the entire unfrozen period of summer (Mastepanov et al., 2008).

Halocarbons

Industrial halocarbons such as the chlorofluorocarbons CFC-10 (CCl_4), CFC-11 (CFCl_3), and CFC-12 (CF_2Cl_2) have been used widely as aerosol propellants, in refrigeration, and in a variety of manufacturing processes. As shown in Fig. 1.25, the production of these gases increased sharply after World War II. Their rate of increase abated after 1970, in the wake of heightened environmental concern and reduced demand. Following the Montreal Protocol, which led to a multilateral ban on CFCs, their production ground to a halt in the mid 1990s.

These anthropogenic species are stable in the troposphere, where their water-insolubility makes them immune to normal scavenging processes associated with precipitation. Because they are long-lived, CFCs are well-mixed in the troposphere, with CFC-11 and CFC-12 having nearly uniform mixing ratios there of $r_{\text{CFC-11}} \cong 0.2$ ppbv and $r_{\text{CFC-12}} \cong 0.3$ ppbv, respectively.

The greatest interest in CFCs surrounds their impact on the ozone layer. Since they are long-lived in the troposphere, chlorofluorocarbons are eventually transported into the stratosphere. There, they are photo-dissociated by UV radiation. Although uniform in the troposphere, mixing ratios of CFC-11 and -12 therefore decrease in the stratosphere (Fig. 1.23). Free chlorine that is released via dissociation of CFCs can destroy ozone. So can free bromine, which is released in similar fashion. Reservoir species for these free radicals are collected in the Equivalent

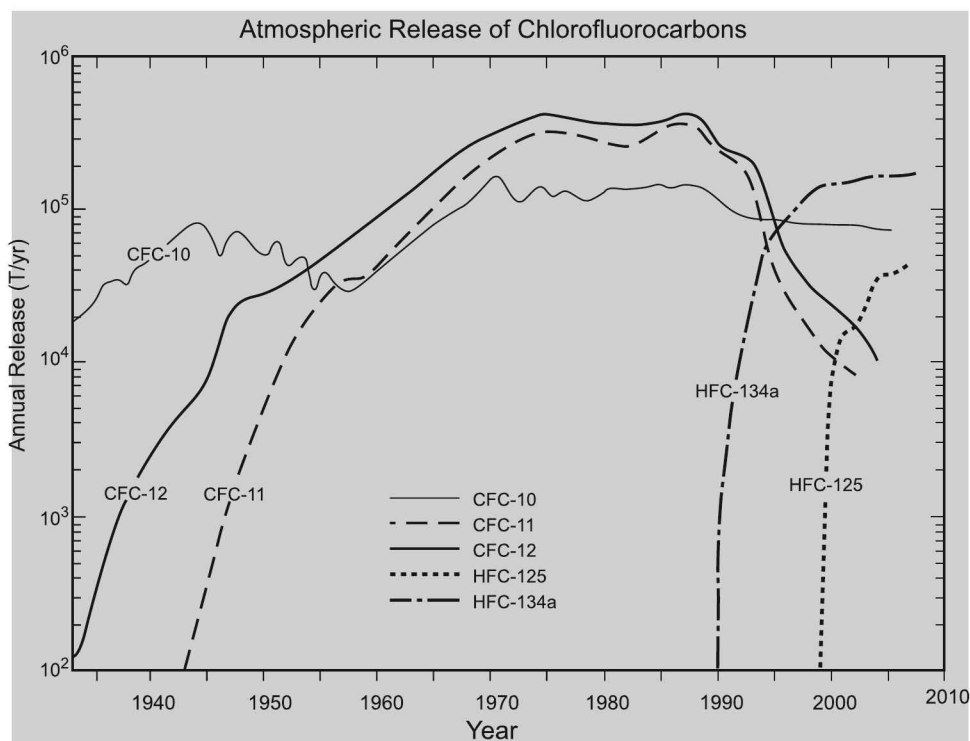


Figure 1.25 Production and eventual emission of chlorofluorocarbons (CFCs) and hydrofluorocarbons (HFCs). Sources: <http://www.afeas.org>, WMO (2006).

Effective Stratospheric Chlorine (EESC). EESC increased during the 1970s and 1980s – by more than 100% (see Fig. 18.36). After peaking in the mid 1990s, EESC began a gradual decline in which values have decreased by about 15%.

Replacing CFCs are hydrofluorocarbons, such as HFC-134a (CF_3CFH_2) and HFC-125 ($\text{CF}_3\text{CF}_2\text{H}$). HFCs have lifetimes in the troposphere that are shorter than those of CFCs. For this reason, HFCs suffer greater removal before they can reach the stratosphere. Nonetheless, like their predecessors, HFCs are steadily increasing (Fig. 1.25).

Photodissociation of CFCs in the stratosphere produces free radicals of chlorine and bromine, which are then said to be “activated.” Through a series of reactions, Cl and Br support catalytic destruction of O_3 , which leads to the formation of the Antarctic ozone hole in Fig. 1.22b. Those free radicals are but one of several ingredients involved in the formation of the ozone hole. Very high cloud in the stratosphere, which rarely forms except over Antarctica, also plays a key role. So does the stratospheric circulation, which controls temperature and the spring breakdown of the polar-night vortex, when the ozone hole disappears.

Other concerns regarding CFCs and HFCs surround their involvement in the Earth’s energy budget. Like CO_2 and O_3 , industrial halocarbons interact with

radiation. Consequently, they too are greenhouse gases. Collectively, their contribution to increased warming of the Earth's surface is thought to be comparable to that of methane (Sec. 8.7).

Nitrogen compounds

Oxides of nitrogen, such as nitrous oxide (N_2O) and nitric oxide (NO), are also relevant to the photochemistry of ozone. Nitrous oxide is produced primarily by natural means relating to bacterial processes in soils. Anthropogenic sources of N_2O include nitrogen fertilizers, combustion of fossil fuel, and biomass destruction. They may account for as much as 25% of its total production. These sources have altered the natural cycle of nitrogen by introducing it in the form of N_2O instead of N_2 . Like methane, N_2O is long-lived and therefore well-mixed in the troposphere. There, it has a nearly uniform mixing ratio of $r_{N_2O} \cong 300$ ppbv (Fig. 1.23).

In the stratosphere, r_{N_2O} decreases with altitude because of dissociation of nitrous oxide. That, in turn, represents the primary source of stratospheric NO. Like free chlorine, NO can destroy ozone catalytically. Nitric oxide is also produced as a by-product of inefficient combustion (e.g., in aircraft exhaust). Nitrous oxide has increased steadily in recent years, similar to other gases with anthropogenic sources. Beyond its relevance to ozone, N_2O is also a greenhouse gas. Its contribution to increased warming of the Earth's surface is thought to be about half as large as that of methane.

Atmospheric aerosol

Suspensions of liquid and solid particles are relevant to radiative as well as chemical processes. Aerosol particles range in size from thousandths of a micron to several microns. They promote cloud formation, making them vital to atmospheric behavior. Aerosol particles serve as condensation nuclei for water droplets and ice crystals, which do not form readily in their absence. These small particulates are produced naturally (e.g., as dust, sea salt, and volcanic debris). They are also produced anthropogenically through combustion and industrial processes. An important source of aerosol is *gas-to-particle conversion*. It occurs through chemical reactions involving, among other precursors, the gaseous emission sulfur dioxide (SO_2). Sulfur dioxide is produced by industry and naturally by volcanos.

Aerosol concentrations are high in urban areas and near industrial complexes (Fig. 1.26). Reflecting anthropogenic sources, their concentrations there are as much as an order of magnitude greater than over maritime regions. Even higher concentrations accompany windblown silicates during desert dust storms (cf. Figs. 9.43; 9.44) and smoke from forest fires and biomass burning (Cover). In addition to their role in cloud formation, aerosol particles interact with radiation: They scatter solar radiation at visible wavelengths. They absorb IR radiation that is emitted by the Earth's surface and atmosphere. These radiative properties involve aerosol in the global energy balance, where it has a role analogous to greenhouse gases. In that capacity, its influence is thought to be as great as that of CO_2 (Fig. 8.30).

Aerosol also plays a key role in chemical processes. One of the more notable is heterogeneous chemistry, which requires the presence of multiple phases. Reactions

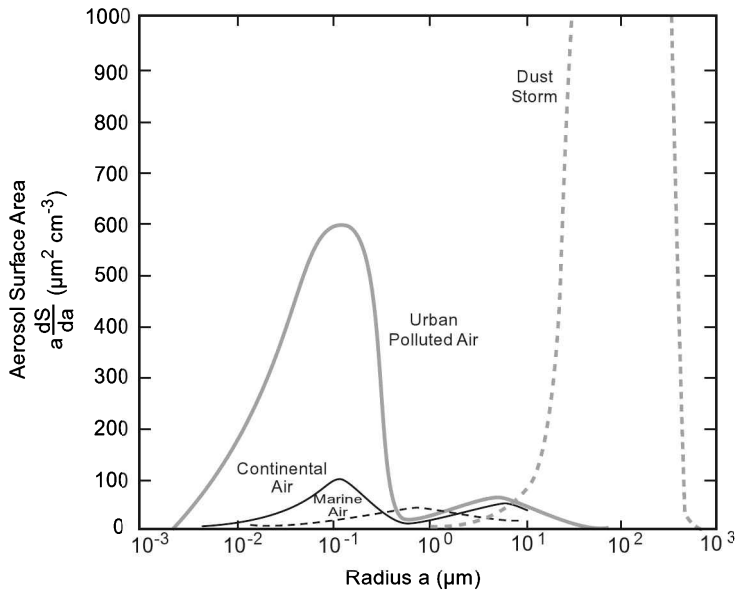


Figure 1.26 Size spectrum of aerosol surface area, as a function of particle radius. The representation is area-preserving: equal areas under the curves represent equal number densities of aerosol particles. *Source:* Slinn (1975).

involving chlorine take place on the surfaces of cloud particles. They lie at the heart of the chain of events that culminates in the formation of the Antarctic ozone hole (Fig. 1.22b).

Beyond this role, aerosol is thought to figure in fluctuations of climate, possibly in the evolution of the atmosphere. Sporadic increases of aerosol produced by major volcanic eruptions have been linked to changes in thermal, optical, and chemical properties of the atmosphere. Figure 1.27a plots the change of global-mean temperature in the lower troposphere following the eruption of Mt. Pinatubo in 1991. Pinatubo introduced volcanic gas and debris well into the stratosphere. Its enhancement of aerosol there remained long enough to modify transfers of radiative energy, in particular, to diminish solar radiation that reaches the Earth's surface. Global-mean temperature decreased following the eruption, attaining a maximum depression of ~ 0.5 K one and a half years later. The depression remained for another 2 years, while global-mean temperature gradually returned to its unperturbed value. Accompanying the depression of temperature was a depression of the column abundance of water vapor (Fig. 1.27b). Precipitable water vapor decreased in much the same fashion, attaining a maximum depression at about the same time of ~ 0.5 mm. Much of the reduction in the column abundance of water vapor derived from humidity changes in the middle and upper troposphere. Their resemblance to the evolution of temperature in the lower troposphere points to a modification of the mechanism by which the troposphere is humidified. Involving convection, such behavior is developed in Chaps. 5 and 7.

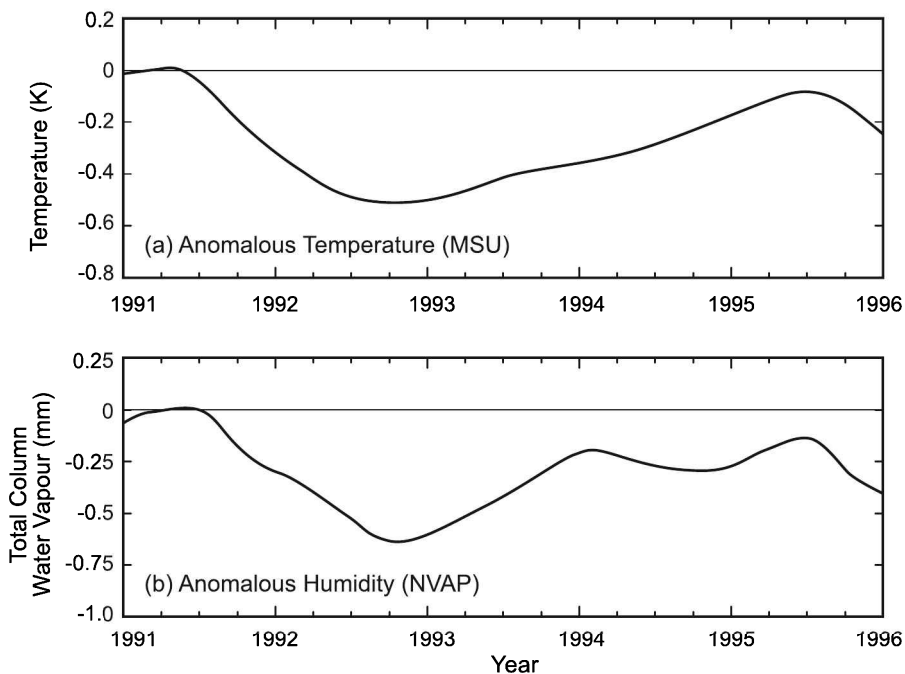


Figure 1.27 (a) Anomalous temperature observed by the Microwave Sounding Unit (MSU) and (b) anomalous column abundance of water vapor from the NASA Water Vapor Project (NVAP) following the eruption of Pinatubo. *Source:* Soden et al. (2002).

1.2.5 Cloud

One of the most striking features of the Earth, when viewed from space, is its extensive coverage by cloud. At any instant, about half of the planet is cloud-covered. Cloud appears with a wide range of shapes, sizes, and microphysical properties. Like trace species, the cloud field is highly dynamic because of its dependence on the circulation. Beyond these more obvious characteristics, cloud and related phenomena figure prominently in a variety of atmospheric processes.

The reason cloud is so striking in satellite imagery is that it reflects back to space a large fraction of incident solar radiation, which is concentrated at visible wavelengths. Owing to its high reflectance at visible wavelengths, cloud shields the planet from solar radiation. This is one important role that cloud plays in the Earth's energy budget. Cloud also plays a major role in the budget of terrestrial radiation, which is emitted in the IR by the Earth's surface and atmosphere to offset solar heating. Because particles of water and ice absorb strongly in the IR, cloud increases the atmosphere's opacity to terrestrial radiation and hence its trapping of radiant energy near the Earth's surface.

Its optical properties enable cloud to be observed from space in measurements of visible and IR radiation. Figure 1.28 shows the IR image from the 11- μm channel of Meteosat-2 that is contemporaneous with the water vapor image in Fig. 1.18. The gray scale in Fig. 1.28 emphasizes the highest objects emitting in the IR. Bright



Figure 1.28 Infrared image at 1200 UT on March 4, 1984, from the $11\text{-}\mu\text{m}$ channel of Meteosat-2. Gray scale displays equivalent blackbody temperature from warmest (black) to coldest (white). (See Fig. 1.29 for geographical landmarks). Supplied by the European Space Agency.

areas correspond to cold high cloud, whereas dark areas indicate warm surface under cloud-free conditions. Many features in the IR image are well correlated with features in the water vapor image. This is especially true of high convective cloud in the tropics, which extends upward to the tropopause. Having horizontal dimensions of tens to a few hundred kilometers, that cloud marks deep convective tower that displaces surface air upward on time scales of hours. On the other hand, the IR image clearly reveals the subtropical Sahara and Kalahari deserts in dark areas that, in the water vapor image, are masked.

High cloud at mid-latitudes also corresponds well with features in the water vapor image. Most striking is the band of high cirrus that extends northeastward from South America, across the Atlantic, and into Africa (cf. Fig. 1.29 for geographical landmarks). Coinciding with the plume of high humidity in Fig. 1.18, this high cloud cover is part of a frontal system that accompanies the cyclone in the eastern Atlantic (cf. Fig. 1.9a). In Figs. 1.18 and 1.28, the cyclone appears as a spiral of water vapor and low cloud. As it advances, the cyclone draws humid tropical air poleward ahead of the front, entraining it with drier mid-latitude air that has been drawn equatorward behind the front.

Cloud also appears in visible imagery, like that shown in Fig. 1.29, which is contemporaneous with the IR and water vapor images. Unlike the latter two images, which



Figure 1.29 Visible image at 1200 UT on March 4, 1984, from Meteosat-2. Supplied by the European Space Agency.

follow from emission of IR radiation, the visible image represents reflected solar radiation. For this reason, it does not depend on the temperature of individual features. Low and comparatively warm stratiform cloud therefore appears just as prominently as high and cold convective cloud. The prevalence of stratiform features in Fig. 1.29 indicates that even shallow cloud is an efficient reflector of solar radiation. Its extensive coverage of the globe makes such cloud an important consideration in the Earth's radiation budget. Swirls of light and dark in the water vapor image, which are signatures of horizontal motion, are only weakly evident in the IR image. They are completely absent in the visible image.

The roles cloud plays in the budgets of solar and terrestrial radiation make it a key ingredient of climate. In fact, the influence cloud cover exerts on the Earth's energy balance is an order of magnitude greater than that of CO_2 . Its dependence on the circulation, thermal structure, and distribution of water vapor make cloud an especially interactive component of the climate system.

With the exception of shallow stratus, most cloud in Figs. 1.28 and 1.29 develops through vertical motion. Two forms of convection are distinguished in the atmosphere. *Cumulus convection* is often implied by the term convection alone. It involves thermally-driven circulations that operate on horizontal dimensions of order 100 km and smaller. Deep tropical cloud in Fig. 1.28 is a signature of cumulus convection. It displaces surface air upward on small horizontal dimensions. *Sloping convection* is

associated with forced lifting, when one body of air overrides another. It occurs coherently over large horizontal dimensions. The band of high cloud preceding the cyclone in the eastern Atlantic is a signature of sloping convection.

Beyond its involvement in radiative processes, convection plays a key role in the dynamics of the atmosphere and in its interaction with the ocean. Deep convection in the tropics releases large quantities of latent heat when water vapor condenses and precipitates back to the Earth's surface. Derived from heat exchange with the ocean, latent heating in the tropics represents a major source of energy for the atmosphere. For this reason, deep cumulus cloud is often considered as a proxy for atmospheric heating.

Figure 1.30a shows a nearly instantaneous image of the global cloud field, as constructed from 11- μm radiances measured aboard six satellites. The highest (brightest) cloud is found in the tropics, inside a narrow band of cumulus convection. Marking the *Inter-Tropical Convergence Zone* (ITCZ), this band of organized convection is oriented parallel to the equator. Exceptional are the tropical landmasses: South America, Africa, and the *maritime continent* of Indonesia and the surrounding archipelago, where the zone of convection widens. Inside the ITCZ, deep convection is supported by the release of latent heat when moisture condenses. Vertical transfers of moisture and energy make the ITCZ important to the tropical circulation and to interactions between the atmosphere and ocean.

The ITCZ emerges conspicuously in the time-mean cloud field, shown in Fig. 1.30b. Over maritime regions, the time-mean ITCZ appears as a narrow strip parallel to the equator. It reflects the convergence of surface air from the two hemispheres inside the Hadley circulation (see Fig. 1.35). Over tropical landmasses, time-mean cloud cover expands due to the additional influence of surface heating, which triggers convection diurnally. There, as throughout the tropics, the unsteady component of the cloud field is as large as the time-mean component. In addition to varying on short time scales, tropical convection migrates north and south annually with the sun. Its annual swing culminates during the solstices in the monsoons over southeast Asia and northern Australia.

Time-mean cloud also reveals the North Atlantic and North Pacific storm tracks. Coinciding with intensified zonal wind (Fig. 1.9a), they are preferred sectors of longitude, where convection is organized by synoptic weather systems. Individual systems are evident in the instantaneous cloud field in Fig. 1.30a. By contrast, in the Southern Hemisphere, such systems are distributed almost uniformly in longitude, inside an unbroken storm track that circumscribes the Antarctic. The distinction between hemispheres is apparent in Fig. 1.30b: Time-mean cloud at mid-latitudes of the Northern Hemisphere is concentrated in the North Atlantic and North Pacific Oceans. However, at mid-latitudes of the Southern Hemisphere, it is almost circumpolar. This hemispheric asymmetry reflects the relative absence of major orographic features in the Southern Hemisphere, limited to the Andes. In the Northern Hemisphere, which includes the Himalayas, the Alps, and the Rockies, such features are substantial. They excite amplified planetary waves that disturb the zonal circulation, introducing strong longitudinal dependence.

Cloud is also important in chemical processes. Condensation and precipitation constitute the primary removal mechanism for many chemical species. Gaseous pollutants that are water soluble are absorbed in cloud droplets, eventually eliminated when those droplets precipitate to the Earth's surface. Referred to as *rain out*, this mechanism also

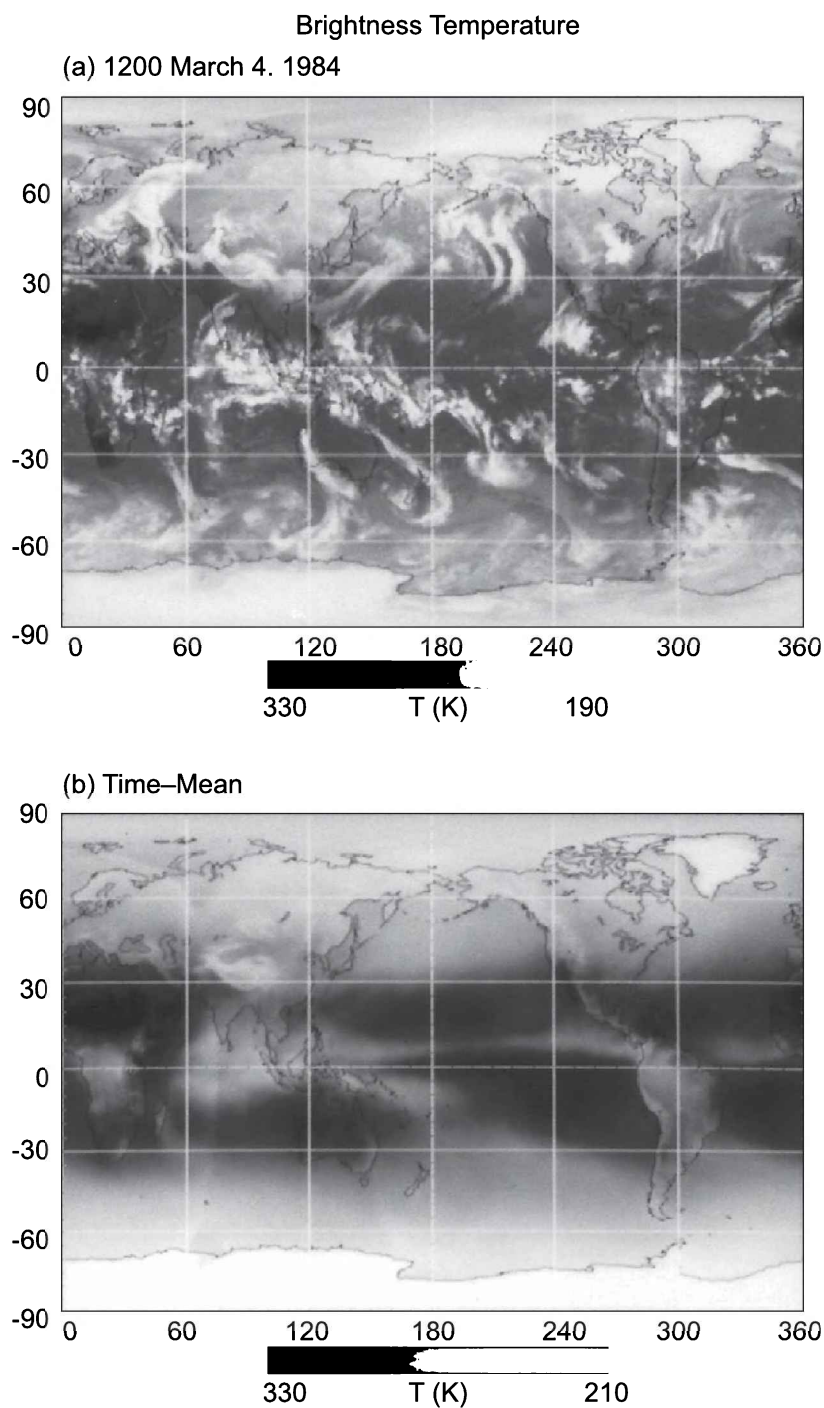


Figure 1.30 Global cloud image constructed from $11\text{-}\mu\text{m}$ radiances measured aboard six satellites simultaneously viewing the Earth. (a) Instantaneous cloud field at 1200 UT on March 4, 1984. (b) Time-mean cloud field for January-March, 1984. Gray scale displays equivalent blackbody temperature as indicated.

scavenges aerosol pollutants, which serve as condensation nuclei for cloud droplets and ice crystals. Although improving air quality, these scavenging mechanisms operate at the expense of surface hydrology. They transfer pollutants to the Earth's surface, often manifest as "acid rain."

Another chemical process in which cloud figures importantly pertains to the ozone hole in Fig. 1.22b. Because moisture is sharply confined to the troposphere, cloud forms in the stratosphere only under exceptionally cold conditions. The Antarctic stratosphere is one of the coldest sites in the atmosphere. As a result, it is populated by a rare form of cloud. Tenuous and very high, Polar Stratospheric Cloud (PSC) is common over the Antarctic during winter. On the surface of PSC, particles operate heterogeneous chemical reactions that involve chlorine and bromine (Sec. 1.2.4). They figure centrally in the formation of the ozone hole each year during Austral spring.

1.3 RADIATIVE EQUILIBRIUM OF THE EARTH

The driving force for the atmosphere is the absorption of solar energy at the Earth's surface. Over time scales long compared with those involved in the redistribution of energy, the Earth-atmosphere system is in thermal equilibrium. The net energy gained must then vanish. Solar radiation is concentrated at visible wavelengths, termed *shortwave (SW) radiation*. Terrestrial radiation, that emitted by the Earth's surface and atmosphere, is concentrated at IR wavelengths, termed *longwave (LW) radiation*. For thermal equilibrium, the absorption of SW radiation must be balanced by emission to space of LW radiation. This basic principle leads to a simple estimate of the mean temperature of the Earth.

The Earth intercepts a beam of SW radiation of cross-sectional area πa^2 and flux F_s (energy/area-time), as illustrated in Fig. 1.31. A fraction of the intercepted radiation, the *albedo* \mathcal{A} , is reflected back to space by the Earth's surface and components of the atmosphere. The remainder of the incident SW flux: $(1 - \mathcal{A})F_s$, is then absorbed by the Earth-atmosphere system. It is distributed across the globe as it spins in the line of the SW beam.

To maintain thermal equilibrium, the Earth and atmosphere must re-emit to space LW radiation at exactly the same rate. Also referred to as *Outgoing Longwave Radiation* (OLR), the emission to space of terrestrial radiation is described by the Stefan-Boltzmann law

$$\pi B = \sigma T^4, \quad (1.29)$$

where πB represents the energy flux integrated over wavelength that is emitted by a blackbody at temperature T and σ is the Stefan-Boltzmann constant. Integrating the emitted LW flux over the Earth and equating the result to the SW energy absorbed obtains the simple energy balance

$$(1 - \mathcal{A})F_s \pi a^2 = 4\pi a^2 \sigma T_e^4, \quad (1.30.1)$$

where T_e is the equivalent blackbody temperature of the Earth. Then

$$T_e = \left[\frac{(1 - \mathcal{A})F_s}{4\sigma} \right]^{\frac{1}{4}} \quad (1.30.2)$$

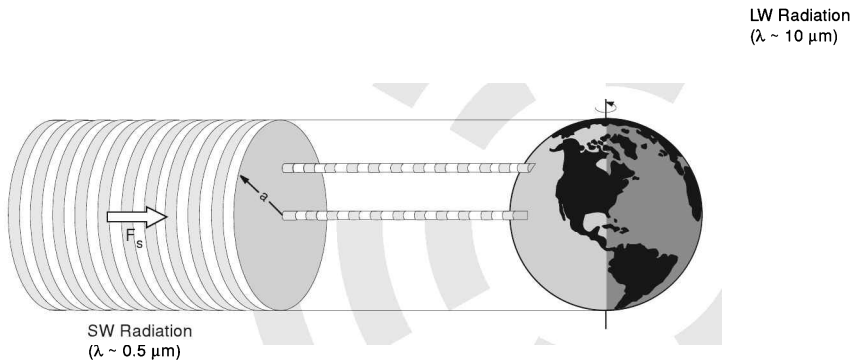


Figure 1.31 Schematic of SW radiation intercepted by the Earth and LW radiation emitted by it.

provides a simple estimate of the Earth's temperature. An incident SW flux of $F_s = 1372 \text{ W m}^{-2}$ and an albedo of $\mathcal{A} = 0.30$ lead to an equivalent blackbody temperature for Earth of $T_e = 255 \text{ K}$. This value is some 30 K colder than the global-mean surface temperature, $T_s = 288 \text{ K}$ (Fig 1.2).

The discrepancy between T_s and T_e follows from the different ways the atmosphere processes SW and LW radiation. Although nearly transparent to SW radiation (wavelengths $\lambda \sim 0.5 \mu\text{m}$), the atmosphere is almost opaque to LW radiation ($\lambda \sim 10 \mu\text{m}$) that is re-emitted by the Earth's surface. For this reason, SW radiation passes relatively freely to the Earth's surface, where it can be absorbed. However, LW radiation emitted by the Earth's surface is captured by the overlying air, chiefly by the major LW absorbers: water vapor and cloud. Energy absorbed in an atmospheric layer is re-emitted, half upward and half back downward. The upwelling re-emitted radiation is absorbed again in overlying layers, which subsequently re-emit that energy in similar fashion. This process is repeated until LW energy is eventually radiated beyond all absorbing components of the atmosphere and rejected to space. By inhibiting the transfer of energy from the Earth's surface, repeated absorption and emission by intermediate layers of the atmosphere traps LW energy, elevating surface temperature over what it would be in the absence of an atmosphere.

The elevation of surface temperature that results from the atmosphere's different transmission characteristics to SW and LW radiation is known as the *greenhouse effect*. The greenhouse effect is controlled by the IR opacity of atmospheric constituents,

which radiatively insulate the Earth's surface. In the atmosphere, the primary absorbers are water vapor and cloud. Carbon dioxide, ozone, methane, and nitrous oxide are also radiatively active at wavelengths of LW radiation, as are aerosol and halo-carbons.

1.4 THE GLOBAL ENERGY BUDGET

Because it follows from a simple energy balance, the equivalent blackbody temperature of the Earth provides some insight into where LW radiation is ultimately emitted to space. The value $T_e = 255$ K corresponds to the middle troposphere, above most of the water vapor and cloud. Most of the energy received by the atmosphere is supplied from the Earth's surface, where SW radiation is absorbed. Transfers of energy from the surface constitute a heat source for the atmosphere. Conversely, LW emission to space by the middle troposphere constitutes a heat sink for the atmosphere. Representing heating and cooling, these energy transfers drive the atmosphere into motion. It then functions as a global heat engine in the energy budget of the Earth, as described in Chap. 3.

Radiative energy absorbed at the Earth's surface must be transmitted to the middle troposphere, where it is rejected to space as LW radiation. From the time it is absorbed at the surface as SW radiation until it is eventually rejected to space, energy assumes a variety of forms. Most of the energy transfer between the Earth's surface and the atmosphere occurs through LW radiation. In addition, energy is transferred through thermal conduction, which is referred to as the transfer of *sensible heat*. It is also transferred through *latent heat*, when water vapor that is absorbed by the atmosphere condenses, releasing its latent heat of vaporization, and then the resulting condensate precipitates back to the Earth's surface.

In addition, energy is stored internally in the atmosphere, in thermal and mechanical forms. The thermal or internal energy of air is measured by temperature. It represents the random motion of molecules. Mechanical energy is represented in the distribution of atmospheric mass within the Earth's gravitational field (potential energy), as well as in the motion of air (kinetic energy). These mechanical forms of energy are involved in the redistribution of energy within the atmosphere. However, they do not contribute to globally integrated transfers between the Earth's surface, the atmosphere, and deep space.

1.4.1 Global-mean energy balance

The globally averaged energy budget is illustrated in Fig. 1.32. Incoming solar energy is distributed across the Earth. Therefore, the global-mean SW flux incident on the top of the atmosphere is given by $\bar{F}_s = \frac{F_s}{4} = 343 \text{ W m}^{-2}$, where the factor 4 represents the ratio of the surface area of the Earth to the cross-sectional area of the intercepted beam of SW radiation (Fig. 1.31). Of the incident 343 W m^{-2} , a total of 106 W m^{-2} (approximately 30%) is reflected back to space: 21 W m^{-2} by air, 69 W m^{-2} by cloud, and 16 W m^{-2} by the surface. The remaining 237 W m^{-2} is absorbed in the Earth-atmosphere system. Of this, 68 W m^{-2} (about 20% of the incident SW flux) is absorbed by the atmosphere: 48 W m^{-2} by atmospheric water vapor, ozone, and aerosol and 20 W m^{-2} by cloud. This leaves 169 W m^{-2} to be absorbed by the surface - nearly 50% of that incident on the top of the atmosphere.

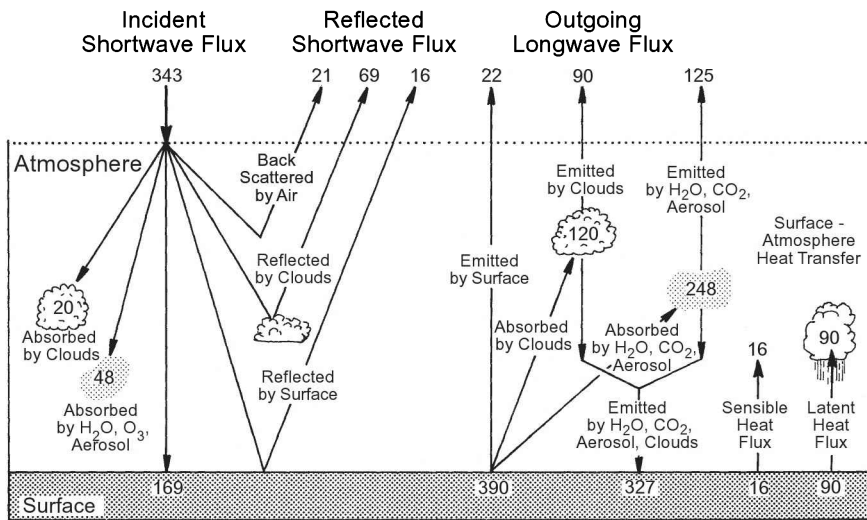


Figure 1.32 Global-Mean Energy Budget ($W m^{-2}$). Updated from *Understanding Climate Change* (1975) with recent satellite measurements of the Earth's radiation budget. Additional sources: Ramanathan (1987), Ramanathan et al. (1989).

The $169 W m^{-2}$ absorbed at the ground must be re-emitted to maintain thermal equilibrium of the Earth's surface. At a global-mean surface temperature of $T_s = 288 K$, the surface emits $390 W m^{-2}$ of LW radiation according to (1.29). This is far more energy than it absorbs as SW radiation. Excess LW emission must be balanced by transfers of energy to the surface from other sources. Owing to the greenhouse effect, the surface also receives LW radiation that is emitted downward by the atmosphere, in the amount $327 W m^{-2}$. Collectively, these contributions result in a net transfer of radiative energy to the Earth's surface:

$$\begin{array}{rccccccc}
 \text{SW Absorption} & + & \text{LW Absorption} & - & \text{LW Emission} & = & \text{Net Radiative Forcing} \\
 & & \text{from Atmosphere} & & & & \text{of Surface} \\
 169 W m^{-2} & + & 327 W m^{-2} & - & 390 W m^{-2} & = & +106 W m^{-2} .
 \end{array}$$

The surplus in absorption of $106 W m^{-2}$ represents net "radiative heating of the surface." For equilibrium, it must be balanced by transfers of sensible and latent heat to the atmosphere. Sensible heat transfer accounts for $16 W m^{-2}$ through conduction with the atmosphere. Latent heat transfer accounts for the remaining $90 W m^{-2}$. It follows from evaporative cooling of the ocean: Heat absorbed from the ocean to evaporate water is transferred to the atmosphere when that water vapor condenses, releases the latent heat of vaporization, and precipitates back to the Earth's surface. Were it not for these auxiliary forms of energy transfer, the Earth's surface would have to be some 50 K warmer for its emission of LW energy to balance its absorption of SW energy.

The atmosphere's energy budget must likewise balance to zero to maintain thermal equilibrium. The atmosphere receives $68 W m^{-2}$ directly through absorption of SW radiation. More important, however, is its absorption of LW radiation emitted by the

Earth's surface. Of the 390 W m^{-2} of LW radiation that is emitted by the Earth's surface, nearly all of it is absorbed by overlying atmosphere: Only 22 W m^{-2} passes freely through the atmosphere and is rejected to space. The remaining 368 W m^{-2} is absorbed by the atmosphere: 120 W m^{-2} by cloud and 248 W m^{-2} by water vapor, CO_2 , and aerosol. The atmosphere loses 327 W m^{-2} through LW emission to the Earth's surface. Another 215 W m^{-2} of LW radiation is rejected to space: 90 W m^{-2} emitted by cloud and 125 W m^{-2} emitted by water vapor, CO_2 , and other minor constituents. Collecting these contributions gives the net transfer of radiative energy to the atmosphere:

$$\begin{array}{rcccccc} \text{SW Absorption} & + & \text{LW Absorption} & - & \text{LW Emission} & - & \text{LW Emission} & = & \text{Net Radiative} \\ & & \text{from Surface} & & \text{to Surface} & & \text{to Space} & & \text{Forcing of} \\ & & & & & & & & \text{Atmosphere} \\ 68 \text{ W m}^{-2} & + & 368 \text{ W m}^{-2} & - & 327 \text{ W m}^{-2} & - & 215 \text{ W m}^{-2} & = & -106 \text{ W m}^{-2}. \end{array}$$

The deficit in absorption of 106 W m^{-2} represents net “radiative cooling of the atmosphere.” That cooling is balanced by mechanical heating: transfers of sensible and latent heat from the Earth's surface.

1.4.2 Horizontal distribution of radiative transfer

The preceding discussion focuses on vertical transfers involved in the global-mean energy balance. Were the absorption of SW energy and the emission of LW energy uniform across the Earth, those vertical transfers could accomplish most of the energy exchange needed to preserve thermal equilibrium. However, geometrical considerations, variations in surface properties, and cloud cover make the horizontal distribution of radiative energy transfer nonuniform.

Even if optical properties of the Earth-atmosphere system were uniform, the absorption of solar energy would not be. At low latitude, SW radiation arrives nearly perpendicular to the Earth (cf. Fig. 1.31). A pencil of radiation incident on the top of the atmosphere is distributed there across a perpendicular cross section. Consequently, the flux of energy crossing the top of the atmosphere at low latitude equals that passing through the pencil. On the other hand, SW radiation at high latitude arrives at an oblique angle. A pencil of radiation incident on the top of the atmosphere is distributed there across an oblique cross section. The latter has area greater than that of a perpendicular cross section. Consequently, the flux of energy crossing the top of the atmosphere at high latitude is less than that passing through the pencil and, therefore, less than the flux arriving at low latitude.

The daily-averaged SW flux incident on the top of the atmosphere defines the *insolation*. Insolation depends on the *solar zenith angle*, which is measured from local vertical. It also depends on the length of day. Both vary with location and season. Figure 1.33 presents the insolation as a function of latitude and month. During equinox, insolation is a maximum on the equator, decreasing poleward like the cosine of latitude. At solstice, the maximum insolation occurs over the summer pole, with values uniformly high across much of the summer hemisphere. In the winter hemisphere, insolation decreases sharply with latitude. It vanishes at the *polar-night terminator*, beyond which the Earth is not illuminated.⁹

⁹ The slight asymmetry between the hemispheres evident in Fig. 1.33 follows from the eccentricity of the Earth's orbit, which brings the Earth closest to the sun during January and farthest from the sun during July.

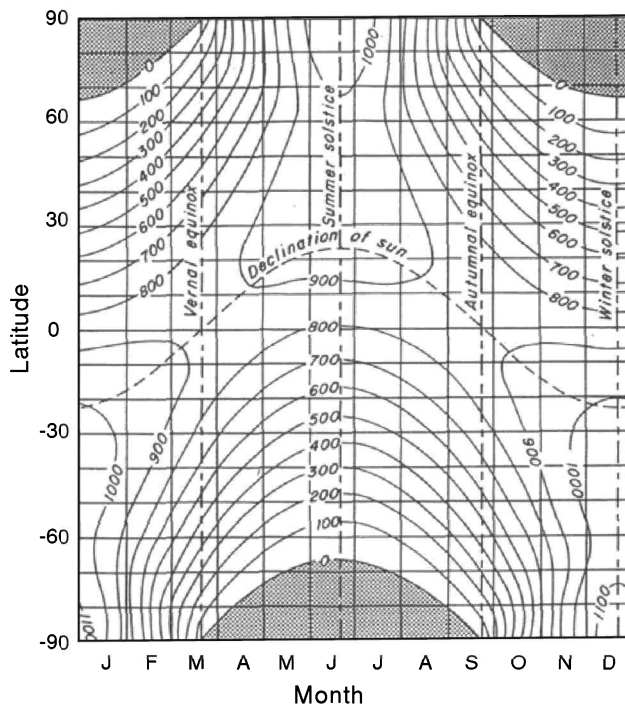


Figure 1.33 Average daily SW flux incident on the top of the atmosphere ($\text{cal cm}^{-2}\text{day}^{-1}$), as a function of latitude and time of year. *Source:* List (1958).

Optical properties of the atmosphere also lead to nonuniform heating of the Earth. At low latitude, SW radiation passes almost vertically through the atmosphere, so the distance traversed through absorbing constituents is minimized. However, at middle and high latitudes, SW radiation traverses the atmosphere along a slant path, which involves a much longer distance and therefore results in greater absorption. A similar effect is introduced through scattering of SW radiation by atmospheric aerosol, which increases sharply with solar zenith angle. Each of these atmospheric optical effects leads to more SW energy reaching the Earth's surface at low latitude than at high latitude.

Optical properties of the Earth's surface introduce similar effects. They emerge in OLR and reflected SW radiation that were measured by the Earth Radiation Budget Experiment (ERBE). The findings of ERBE are collected in Fig. 1.34 for northern winter. Figure 1.34a plots the distribution of albedo. Low latitudes, which account for much of the Earth's surface area, have extensive coverage by ocean. Those regions are characterized by very low albedo and thus absorb most of the SW energy incident on them. High latitudes, particularly in the winter hemisphere, have extensive coverage by snow and ice. Having high albedo, those surfaces reflect back to space much of the SW energy incident on them. This is especially true for large solar zenith angle (intrinsic to high latitudes), which increases scattering. Cloud cover adds to the high albedo at extratropical latitudes, for example, in the storm tracks, where cloud is organized by synoptic weather systems. High albedo is also evident in the tropics, where deep

(a) Albedo

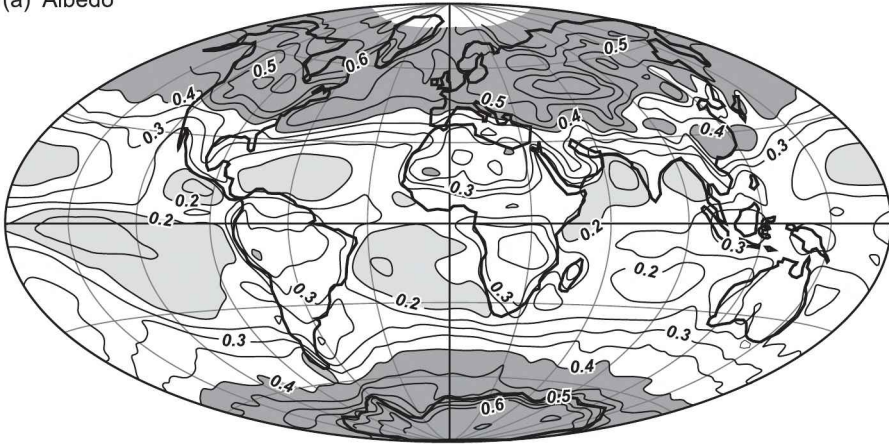
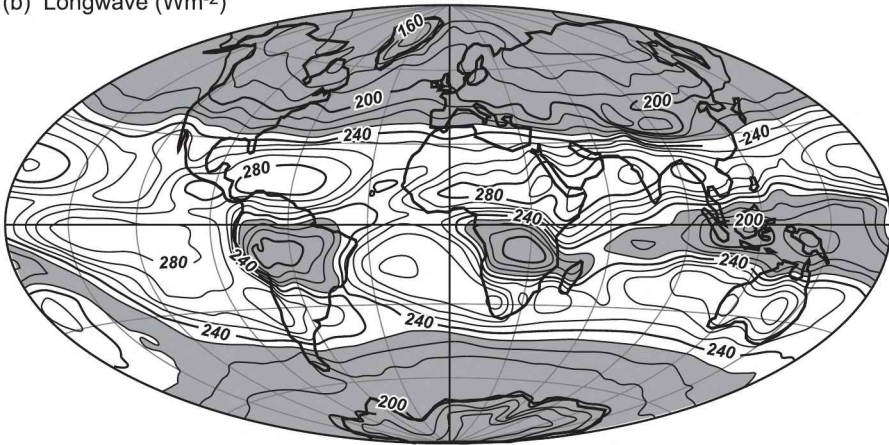
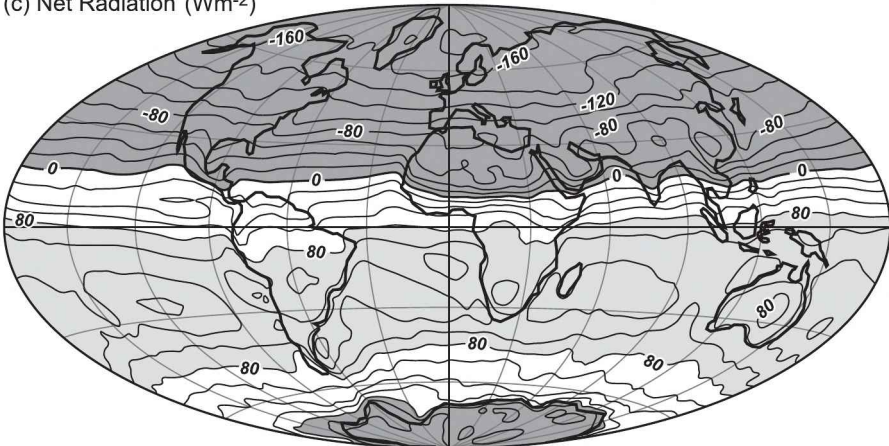
(b) Longwave (Wm^{-2})(c) Net Radiation (Wm^{-2})

Figure 1.34 Top-of-the-atmosphere radiative properties for December-February derived from the Earth Radiation Budget Experiment (ERBE), which was on board the ERBS and NOAA-9 satellites: (a) albedo, (b) outgoing longwave radiation, (c) net radiation. Adapted from Hartmann (1993).

convective cloud leads to maxima over the Amazon Basin, tropical Africa, and Indonesia. Flanking those regions are areas of low albedo. They correspond to cloud-free oceanic regions in the subtropics and between tropical landmasses.

Outgoing longwave radiation (Fig. 1.34b) is dominated by the poleward decrease of temperature, which causes OLR to decrease steadily with latitude (1.29). At middle and high latitudes, anomalously low OLR is found over the Tibetan Plateau and Rocky Mountains, where cold surface temperature reduces LW emission (cf. Fig. 1.30b). Anomalously-low OLR is also found near the equator, over the convective centers of South America, Africa, and Indonesia. High cloud in those regions is very cold. Its LW emission is therefore substantially weaker than emission by warm surface in neighboring cloud-free regions (1.29). In fact, subtropical latitudes are marked by distinct maxima of OLR in regions that flank organized convection. Sinking motion there compensates rising motion inside the ITCZ. It inhibits cloud formation and precipitation, maintaining arid zones that typify the subtropics.

The difference between the SW radiation absorbed by the Earth-atmosphere system and the LW radiation it emits to space defines the *net radiation*. In the global mean, net radiation must vanish for thermal equilibrium (1.30). Locally, however, this need not be the case. The distribution of net radiation observed by ERBE is plotted in Fig. 1.34c. Low latitudes are characterized by a surplus of net radiation: They experience radiative heating. High latitudes, on the other hand, are characterized by a deficit of net radiation: They experience radiative cooling. During northern winter, the entire Northern Hemisphere poleward of 15° absorbs less energy in the form of SW radiation than it emits to space as LW radiation. At the same time, most of the Southern Hemisphere absorbs more energy in the form of SW radiation than it emits to space as LW radiation. Only over Antarctica, where permanent snow cover maintains high albedo, is net radiation negative. Net radiation is positive and large almost uniformly across the tropics and in the subtropics of the summer hemisphere, where strong insolation and cloud-free conditions prevail. Notice that the strong zonal asymmetries that punctuate the SW and LW components individually (Figs. 1.34a,b) are absent. Although individually, albedo and OLR exhibit large geographical variations across the tropics, a cancellation between these components in regions of convection nearly eliminates those anomalies in the distribution of net radiation. Net radiative forcing of the Earth-atmosphere system is therefore almost zonally symmetric.

The optical properties just described lead to nonuniform heating of the Earth-atmosphere system. Low latitudes experience radiative heating, whereas middle and high latitudes experience radiative cooling. To preserve thermal equilibrium, the surplus of radiative energy received by low latitudes must be transferred to middle and high latitudes, where it offsets the deficit of radiative energy. This meridional transfer of energy is accomplished by the general circulation, about 60% of it by the atmosphere.

1.5 THE GENERAL CIRCULATION

The term *general circulation* refers to the aggregate of motions controlling transfers of heat, momentum, and constituents. Broadly speaking, the general circulation also includes interactions with the Earth's surface (e.g., transfers of latent heat and moisture from the ocean) because atmospheric behavior, especially that operating on time scales longer than a month, is influenced importantly by such interactions. For this reason, properties of the Earth's surface, like sea surface temperature (SST), figure importantly in the general circulation of the atmosphere.

The general circulation is maintained against frictional dissipation by a conversion of potential energy, associated with the distribution of atmospheric mass, to kinetic energy, associated with the motion of air. Radiative heating acts to expand the atmospheric column at low latitude, according to hydrostatic equilibrium (1.17). It thus acts to raise the column's center of mass. By contrast, radiative cooling acts to compress the atmospheric column at middle and high latitudes and lower its center of mass. The uneven horizontal distribution of mass that results introduces an imbalance of pressure forces. It drives a meridional overturning, with air rising at low latitude and sinking at middle and high latitudes.

The simple meridional circulation implied in the preceding paragraph is modified importantly by the Earth's rotation. As is evident from the instantaneous circulation at 500 hPa (Fig. 1.9a), the large-scale circulation remains nearly tangential to contours of isobaric height (e.g., tangential to isobars on a constant-height surface). Net radiative heating in Fig. 1.34c tends to establish time-mean thermal structure in which isotherms and contours of isobaric height are oriented parallel to latitude circles. The time-mean circulation at 500 hPa (Fig. 1.10b) is therefore almost westerly and circumpolar at middle and high latitudes. Characterized by a nearly zonal jet stream, the time-mean circulation possesses only a small meridional component to transfer heat between the equator and poles. A similar conclusion applies in the stratosphere, where time-mean motion is strongly zonal (Fig. 1.10b).

Asymmetries in the instantaneous circulation involve meridional deflections of air. Consequently, they play a key role in transferring heat between the equator and poles. In the troposphere, much of the heat transfer is accomplished by unsteady synoptic weather systems. They transport heat poleward through sloping convection, which exchanges cold polar air with warm tropical air. Ubiquitous in the troposphere, those disturbances contain much of the kinetic energy at mid-latitudes. They develop preferentially in the North Pacific and North Atlantic storm tracks and in the continuous storm track of the Southern Hemisphere. By rearranging air, synoptic disturbances also shape the distributions of water vapor and other constituents that are introduced at the Earth's surface.

In the stratosphere and mesosphere, synoptic disturbances are absent (Fig 1.10). Planetary waves, which propagate upward from the troposphere, play a role at these altitudes, similar to the role played by synoptic disturbances in the troposphere. Generated near the Earth's surface, these global-scale disturbances force the middle atmosphere mechanically. By deflecting air across latitude circles, planetary waves transport heat and constituents between low latitudes and high latitudes. Such transport is responsible for the greatest abundances of ozone being found at middle and high latitudes (Fig. 1.21), despite its production at low latitude.

At low latitude, the Earth's rotation exerts a weaker influence on air motion. Kinetic energy there is associated primarily with *thermally direct circulations*, in which air rises in regions of heating and sinks in regions of cooling. Thermally direct circulations in the tropics are forced by the geographical distribution of atmospheric heating (e.g., as implied by the distributions of net radiation in Fig. 1.34c and time-mean cloud in Fig. 1.30b).

Latent heat release inside the ITCZ drives a meridional overturning that prevails within 30° of the equator. Depicted in Fig. 1.35, the *Hadley circulation* involves upwelling near the equator and downwelling at subtropical latitudes. (A meridional section of observed motion is presented in Fig. 15.5.) Further poleward are westerlies, which intensify to form the subtropical jets (Fig. 1.8). At low latitude, equatorward

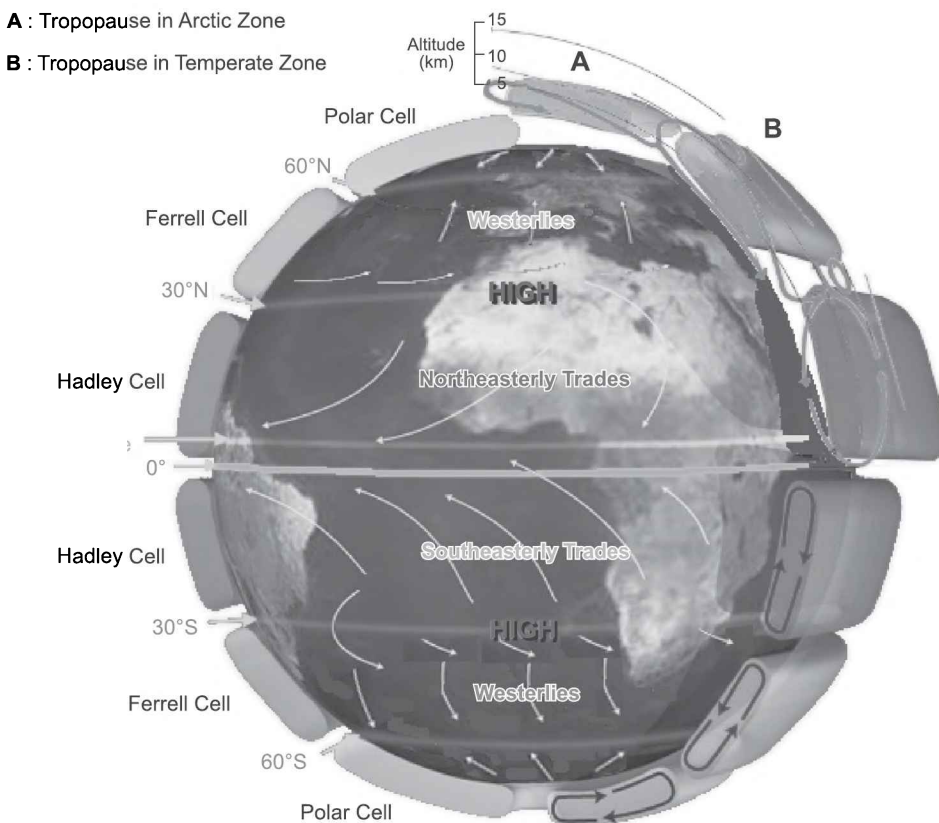


Figure 1.35 Schematic of the mean circulation, which is comprised of Hadley cells equatorward of 30°, Ferrell cells at mid-latitudes, and polar cells at higher latitude. The lower branch of the Hadley cell involves northeasterly and southeasterly Trade Winds, which meet in the InterTropical Convergence Zone (ITCZ), where air rises to form deep cumulus convection and heavy rainfall. *Source*: sealevel.jpl.nasa.gov. See color plate section: Plate 2.

flow near the Earth's surface is manifested by easterly *trade winds*, northeasterly in the Northern Hemisphere and southeasterly in the Southern Hemisphere (cf. Fig. 15.9). They converge into the ITCZ. Upwelling inside the ITCZ is compensated by downwelling at subtropical latitudes. Comprising the descending branch of the Hadley circulation, the latter forms a *subtropical high* in each hemisphere. By suppressing cloud and precipitation, downwelling of the Hadley circulation maintains deserts (cf. Fig. 1.34b). It is for this reason that all major deserts on Earth are found in the subtropics.

Nonuniform heating across the tropics also drives zonal overturning. Known as a *Walker circulation*, the latter involves upwelling at longitudes of heating and downwelling at longitudes of cooling. The nonuniform distribution of land and sea, along with resulting asymmetries in atmospheric heating, lead to Walker circulations along the equator. The concentration of latent heating over the maritime continent (Fig. 1.34b) forces the Pacific Walker circulation, which is illustrated in Fig. 1.36. This circulation reinforces easterly trade winds across the equatorial Pacific. Its descending

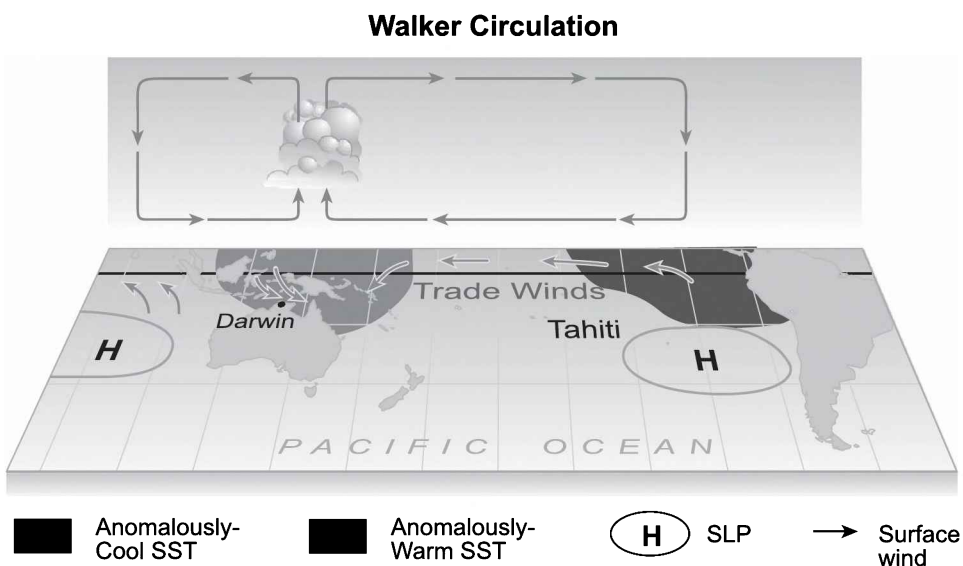


Figure 1.36 Schematic of the Pacific Walker circulation. *Source:* <http://www.bom.gov.au/lam/climate/levelthree/analclim>. See color plate section: Plate 3.

branch, found to the east, maintains the arid climate that typifies the eastern Pacific (cf. Figs. 1.34a,b).

In addition to driving quasi-steady circulations, latent heat release excites unsteady wave motions. Those disturbances are another form of asymmetry in the circulation. Wave motions are also excited mechanically at the Earth's surface. By displacing atmospheric mass vertically, orographic features such as the Alps, the Himalayas, and the Rocky Mountains excite planetary waves and gravity waves that radiate away from those source regions. Like wave activity generated by unsteady heating, orographically-excited disturbances propagate the influence of their forcing to other regions.

Kinetic energy associated with the general circulation is damped by frictional dissipation. Frictional damping occurs principally through turbulence. Involving eddies with a wide range of scales, it homogenizes large-scale gradients. About half of the kinetic energy of the large-scale circulation is dissipated in the lowest kilometer of the atmosphere. Inside the *planetary boundary layer*, small-scale turbulence extracts energy from large-scale motion and cascades it to small dimensions, where it is dissipated by molecular diffusion. In the *free atmosphere*, namely, away from the surface, the large-scale circulation suffers frictional dissipation through convective motion and dynamical instability. Both generate turbulence from large-scale organized motion. The large-scale circulation is also damped by thermal dissipation. LW cooling to space damps atmospheric motion indirectly, by destroying the accompanying temperature anomaly.

1.6 HISTORICAL PERSPECTIVE: GLOBAL-MEAN TEMPERATURE

This chapter closes with an overview of the salient property of climate, global-mean surface temperature, along with evidence of how it has evolved. Surface temperature

dictates many features of climate. Among them is the amount of water that is sequestered in glacial ice. The latter, in turn, influences sea level.

Like any property of climate, Global Mean Temperature (GMT) is comprised of two components: (1) *Climate variability* involves fluctuations that enter through influences that are internal to the climate system, like transient exchanges of heat between the atmosphere and ocean, and influences that are external to the climate system, like variations of solar luminosity. Involving a wide range of time scales (years, decades, and longer), they represent natural fluctuations of the climate system. (2) The *secular* variation of climate evolves systematically in one direction. Unlike climate variability, it is not offset by subsequent swings in the opposite direction. It is in the secular variation that anthropogenic changes should be concentrated.

1.6.1 The instrumental record

Routine measurements of air temperature began in the nineteenth century. They eventually formed a network of ground-based weather observations, which contribute today to the operations of major weather centers to produce short-term forecasts. Beyond surface measurements, the operational network makes vertical soundings of temperature, humidity, and wind via *rawinsonde*. Launched synchronously at 0000 and 1200 Universal Time (UT = local time at Greenwich England), these balloon-borne instruments relay back local conditions during their ascent (nominally reaching 50 hPa). The rawinsonde network, parts of which date back to the 1930s, is comprised of about 1000 stations over the globe, displayed in Fig. 1.37 (open circles). It provides coverage that is dense over populated continents, sparse over remote landmasses like Africa, South America, Greenland, and Antarctica, and almost nonexistent over the oceans. Augmenting the rawinsonde network are numerous other reporting stations, especially for surface measurements, which are superimposed in Fig. 1.37 (dots). Nonetheless, even collectively, the ground-based network suffers from the same nonuniform sampling that limits the rawinsonde network.

Because it is discriminated to continental regions, such sampling can introduce a systematic error or bias into the record of global-mean temperature. Changes over continental regions are inadvertently magnified relative to changes over maritime regions, which are left unaccounted for. The respective error in global-mean temperature, however, appears to be small, a couple of percent – at least as inferred from a climate simulation (Madden and Meehl, 1993).

Exacerbating the sampling error are two related features of continental temperature: Owing to thermal properties of the Earth's surface, diurnal and seasonal temperature extremes are large over land, but small over ocean (Sec. 15.4). The diurnal cycle of temperature, because it is controlled by water vapor and cloud, is not a robust feature of climate simulations (e.g., as have been relied on to evaluate sampling). In addition, many stations in the ground network are, for historical reasons, situated near population centers (Fig. 1.37). Over the course of the instrumental record, those centers expanded through urban development. With their expansion was an amplification of the *urban heat island effect*, wherein urban centers are distinctly warmer than their surroundings.¹⁰ The heat island effect is greatest at night, when rural areas cool faster

¹⁰ The heat island effect develops through several mechanisms, notably through the greater heat capacity of concrete structures and their interruption of LW cooling to space.

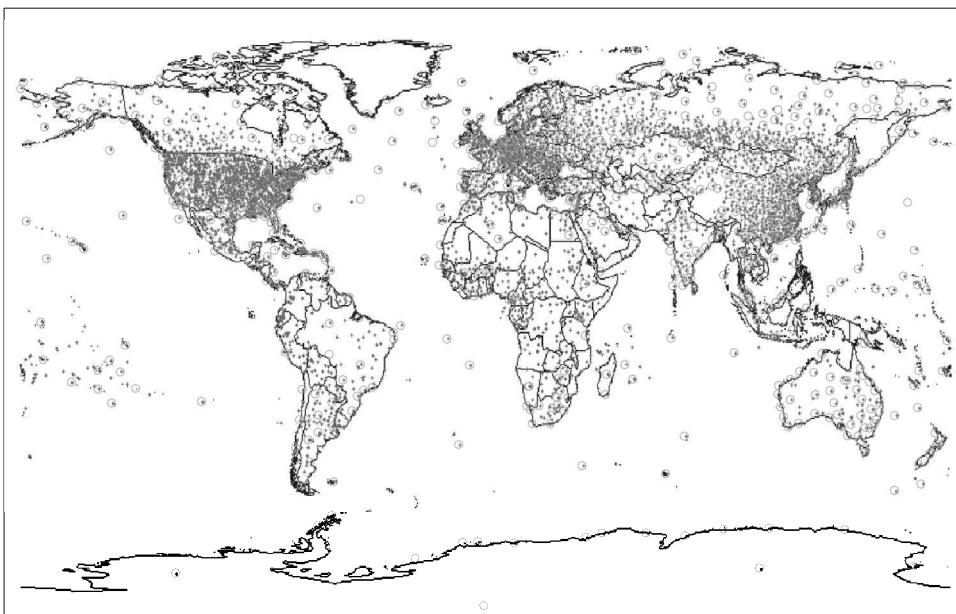


Figure 1.37 Stations comprising the global radiosonde network, where vertical profiles of atmospheric structure are measured twice daily at 0000 and 1200 UT (open circles), and the ground network of surface temperature measurements (dots). *Sources:* Gruber and Haimberger (2008), Peterson and Vose (1997). See color plate section: Plate 4.

and, therefore, achieve colder temperatures than urban centers. It emerges conspicuously in temperature trends.

Figure 1.38a compares records of annual-mean temperature between stations that are in the same region (California) but neighboring disparate populations at the end of the century.¹¹ The average of stations neighboring population centers that exceed 1,000,000 (upper) has a mean over the twentieth century of $\sim 62^{\circ}\text{F}$ (17°C) and a warming trend of some $0.3^{\circ}\text{F}/\text{decade}$. Population centers of less than 100,000 experienced far less growth and urban development. The average of stations neighboring those centers (lower) has a mean over the twentieth century that is some 6°F cooler. It exhibits a warming trend, but one that is an order of magnitude weaker. Hence, more-urbanized regions are warmer and they warmed faster than other regions.

Superimposed in Fig. 1.38a is the record of Global Mean Temperature, derived from the network of surface stations (dotted). The trend exhibited by GMT is closer to that reported by large population centers. The variability of GMT, however, is an order of magnitude weaker than that in the records of regional temperature. Close inspection reveals little correspondence with either. Changes of regional temperature are strongly correlated between neighboring centers of different population. However, they are virtually uncorrelated to changes of GMT. Hence the comparatively large changes that dominate regional climate have little relationship to those of GMT.

The lack of dependence on GMT follows from the spatial coherence of regional changes. Plotted in Fig. 1.38b is the same information for SE Australia, a region at

¹¹ Notice that, even in annual means, regional temperature exhibits climate variability of 1–2 K.

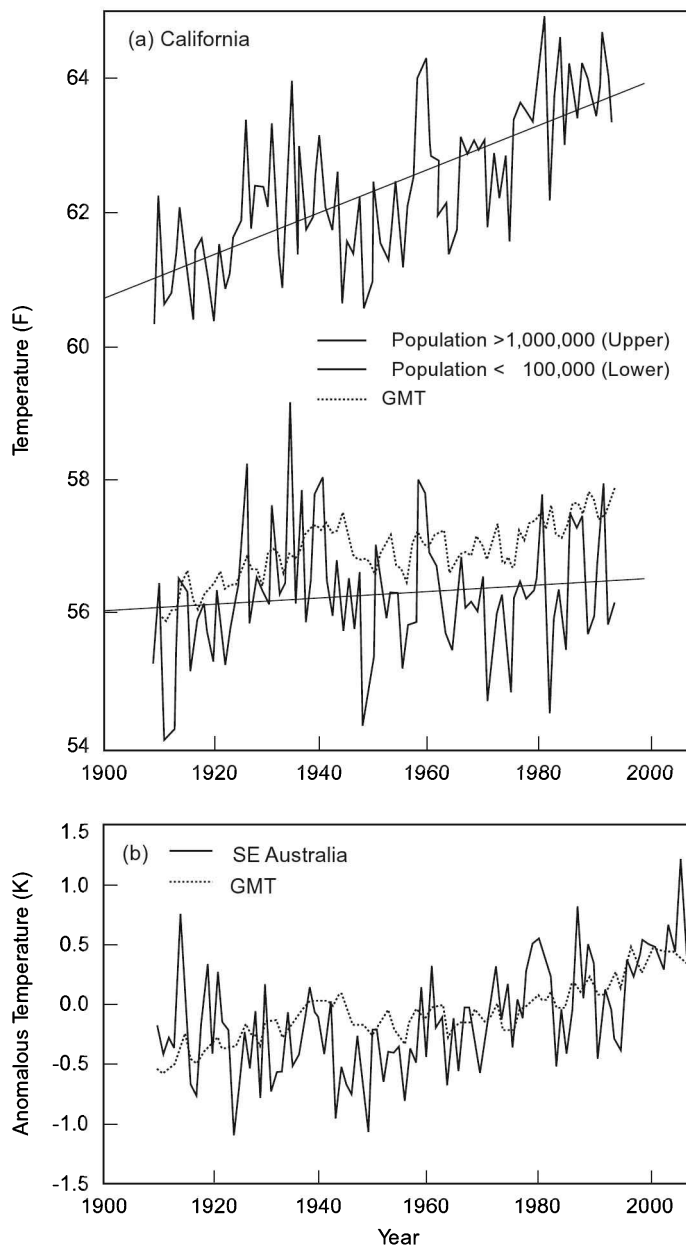


Figure 1.38 (a) Record of annual-mean temperature in California, averaged over population centers exceeding 1,000,000 (upper) and of less than 100,000 (lower). Superimposed is the record of Global Mean Temperature (GMT) from the network of surface stations (dotted). (b) Record of annual-mean temperature over SE Australia (undiscriminated by population), a region at conjugate latitude, of comparable area, and proportionate population growth (solid), and GMT (dotted). (c) Histogram of observed temperature trend over California, as a function of population. *Sources:* Goodridge (1996); Robinson et al. (1998).

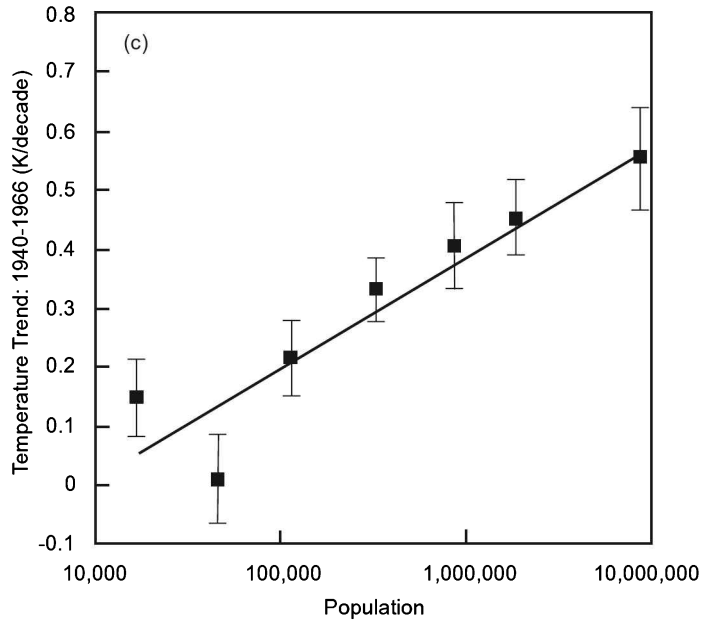


Figure 1.38 (*continued*)

conjugate latitude, of comparable area, and proportionate population growth. The picture (undiscriminated by population) is much the same: Even in annual means, changes of regional temperature exhibit little correspondence to changes of GMT, which are an order of magnitude weaker. The records have a correlation of 0.29. Thus GMT accounts for only about 8% of the variance of regional temperature, rendering one largely independent of the other. Instead, regional temperature is dominated by climate variability, fluctuations that operate on a wide range of time scales. Included are decadal drifts of regional climate. For example, during the first half of the twentieth century, temperature over SE Australia declined, whereas GMT then increased. During the last two decades of the twentieth century, both increased. Afterward, the records diverged again, GMT no longer increasing.

The correlation between changes over SE Australia and California is even smaller, 0.24. Only about 6% of the variance operates coherently between those regions. It follows that changes of regional temperature, even in annual means, have limited spatial coherence. When averaged over the Earth, those spatially incoherent changes cancel. Left behind are much weaker changes that are globally coherent and therefore contribute to changes of GMT. Changes of regional temperature that operate on shorter time scales, for example, represented in seasonal means, exhibit even greater variance, with even less dependence on GMT. So do episodes of extreme weather, which develop through anomalous regional conditions. As is apparent in Figs. 1.38a and 1.38b, such conditions are dominated by climate variability, which is largely independent of GMT.

The long-term evolution of station reports in Fig. 1.38a suggests a correspondence to population growth. That relationship is borne out more generally. Plotted in Fig. 1.38c is a histogram of the temperature trend reported by stations, as a function of neighboring population at the end of the twentieth century. The reported trend

is strongly correlated with population in the surroundings. Stations in large urban centers that underwent major development report magnified trends. They are unrepresentative of trends from smaller population centers in the wider surroundings. The distorted contribution to global-mean temperature is intrinsic to populated landmasses, which are sampled by the ground network (Fig. 1.37). It is not representative of remote landmasses and maritime regions, which are left unsampled.

Circumventing the sampling limitations of the ground network are satellite observations. The satellite era began around 1979, when atmospheric structure and composition began to be monitored from space by several operational and research platforms. Most are in a polar orbit, which, on time scales longer than a day, provides uniform and nearly complete coverage of the Earth. However, the benefit of satellite observations in horizontal sampling is offset by their limitation in vertical sampling. Unlike surface measurements and rawinsondes, which provide temperature in the lowest 100 m, satellite retrievals describe temperature that is averaged over a comparatively deep layer of the atmosphere, characterizing the lowest 5 km. Accordingly, sharp horizontal changes of surface temperature (e.g., associated with abrupt differences between land and ocean) are lost, blended with more gradual horizontal variations that prevail at upper levels. Nonetheless, the coverage and continuity afforded by the satellite record makes it unrivalled in considerations of global-mean temperature.

Plotted in Fig. 1.39, by month, is the satellite record of GMT (dashed) recovered from measurements of channel 2 of the Microwave Sounding Unit (MSU) and channel 5 of the Advanced Microwave Sounding Unit (AMSU). Those measurements represent temperature in the lower troposphere. Superimposed is the corresponding record derived from surface measurements (solid). The two records exhibit much the same behavior, deviating only in detail. Both evidence an upward trend. GMT warmed since 1982, by some 0.19 K/decade in the surface record and by 0.16 K/decade in the satellite record. About the trend are interannual fluctuations of order ± 0.20 K. They represent the climate variability of GMT. Conspicuous among interannual fluctuations is the sharp warming during 1997–1998. It coincides with the El Niño during that period (Chap. 17). El Niño is an oceanic phenomenon. Its signature is only half as strong in the surface record of GMT, which is discriminated to continent, as it is in the satellite record, which has uniform coverage of the Earth. A signature of the 1982–1983 El Niño is less evident. Nevertheless, it exhibits a similar discrepancy between the surface and satellite records of GMT. Also evident is influence from the eruption of Pinatubo in 1991 (cf. Fig. 1.27). After declining for about 2 years, global temperature rebounded, increasing for about a decade. Global-mean temperature then leveled off around 2002, followed by several years of general decline.

In the surface record, much of the warming trend derives from continental regions (Fig. 1.40a); that is, after all, where those observations are concentrated (cf. Fig. 1.38). It is especially strong at polar sites of the Northern Hemisphere, which are glaciated during winter. The trend over continental regions follows principally from nighttime temperature, which exhibits faster warming than daytime temperature (Easterling et al., 1997). The trend derived from the satellite record (Fig. 1.40b) is more uniform longitudinally, with weaker geographical dependence. Nonetheless, it shares major features with the trend in surface record, notably, the strongest warming being found at polar latitudes of the Northern Hemisphere. Offsetting the latter is a cooling trend over the Southern Ocean and Antarctica. Notice that, due to scant coverage, it is virtually unrepresented in the surface record (Fig. 1.40a). Smoother horizontal structure

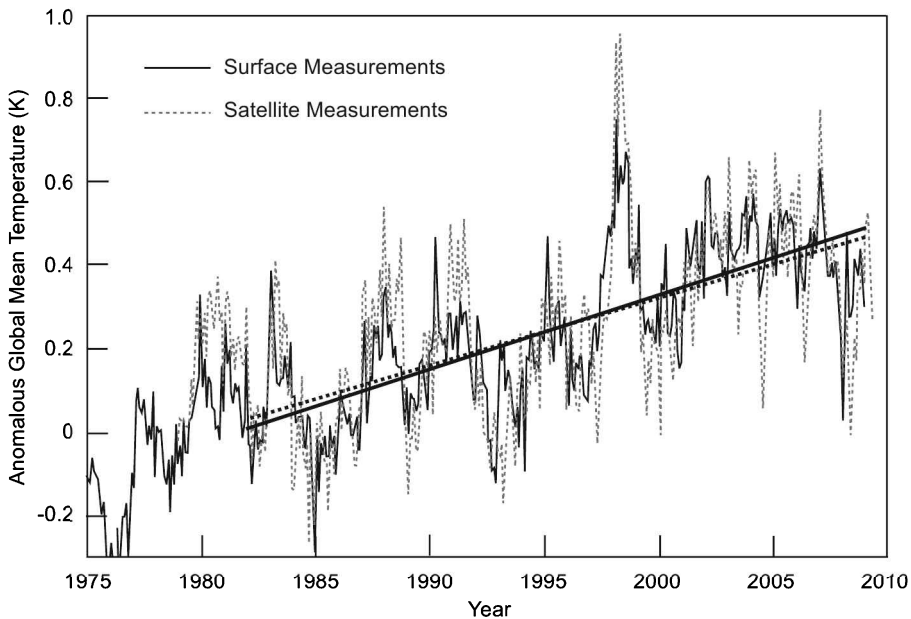


Figure 1.39 Record of Global Mean Temperature (GMT) from station measurements at the Earth's surface (solid) and from satellite measurements by Channel 2 of MSU and Channel 5 of AMSU (dashed). Both evidence a warming trend since 1982, of 0.19 K/decade in the surface record and 0.16 K/decade in the satellite record. (Over 1979–2009, the trend in MSU is ~ 0.125 K/decade.) The satellite record derives from the retrieval by University of Alabama at Huntsville (Christy, Spencer, and Braswell, 2000), in association with National Aeronautics and Space Administration (NASA). A different retrieval of the same measurements has been performed by Remote Sensing Systems (Mears et al., 2003), in association with National Oceanic and Atmospheric Administration (NOAA) – (not shown). It is very similar, albeit with a decadal trend closer to that of the surface record. However, that retrieval omits data from Antarctica, which, during the same interval, exhibits a cooling trend.

intrinsic to the satellite trend reflects the comparatively deep layer that is represented by those measurements. In the vertical average over 5 km, sharp horizontal differences of surface temperature, for example, between land and ocean, are dissolved into structure that is more zonally uniform – smooth structure at upper levels that is maintained by horizontal air motion. Although it blurs geographical features over large horizontal dimension, this process should have little impact on global-mean temperature (Prob. 1.18). It is for this reason that the surface and satellite records of GMT track as closely as they do (Fig. 1.39).

In Fig. 1.41, the ground-based record is extended back to 1850, along with the respective uncertainty. During the latter half of the nineteenth century, GMT meandered, with deviations of a few tenths of a Kelvin. Between 1880 and 1910, it declined by about 0.5 K. Global-mean temperature then increased steadily for three decades, during 1910–1940, also by about 0.5 K. During the 1940s and early 1950s, GMT declined again, followed by two decades when it simply fluctuated. It is noteworthy that the decline

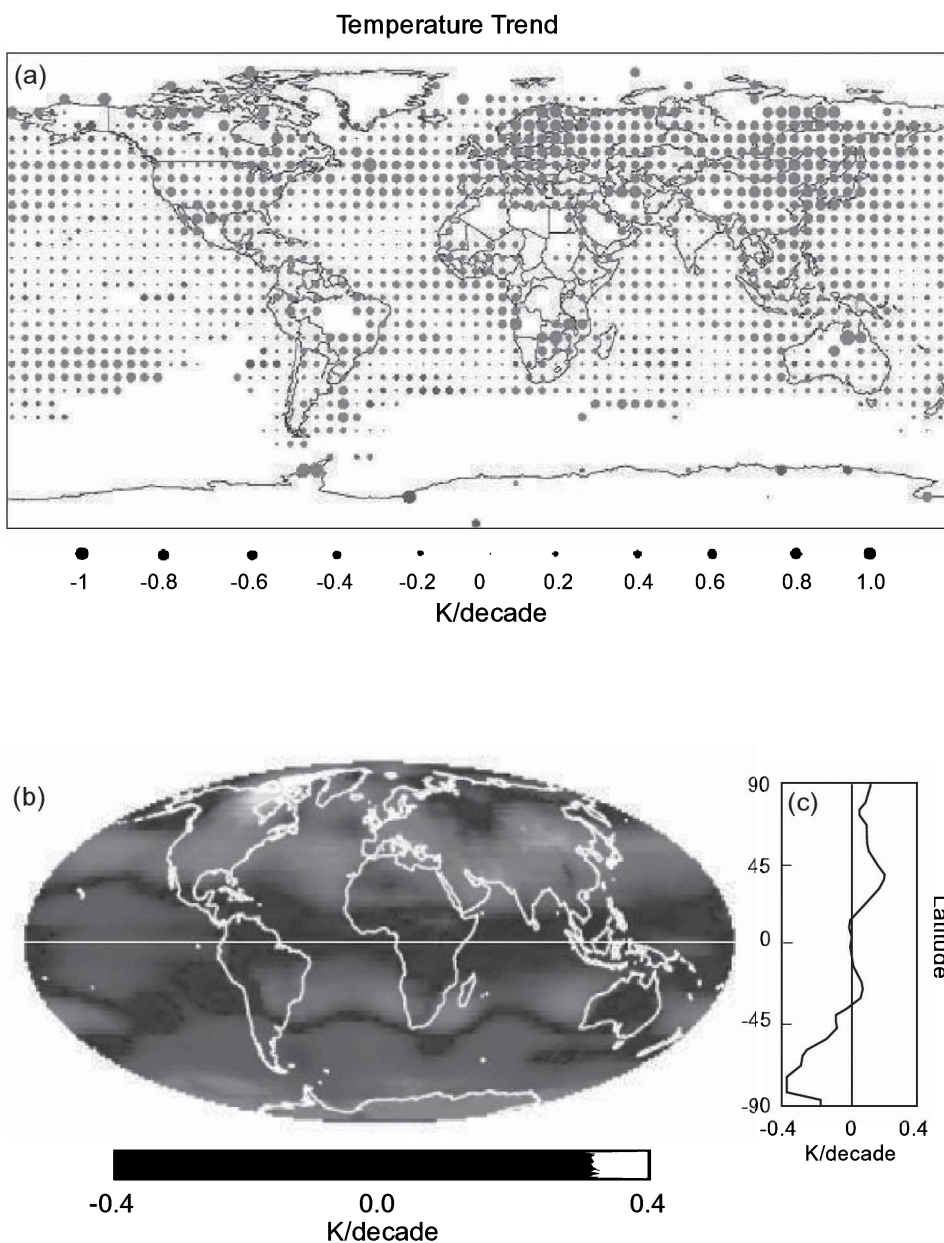


Figure 1.40 Temperature trend during the satellite era, as function of geographical position, derived from (a) the surface record and (b) the satellite record. (c) Zonal-mean trend in the satellite record, which has global coverage. *Sources:* IPCC (2007), Mears et al. (2003). See color plate section: Plate 5.

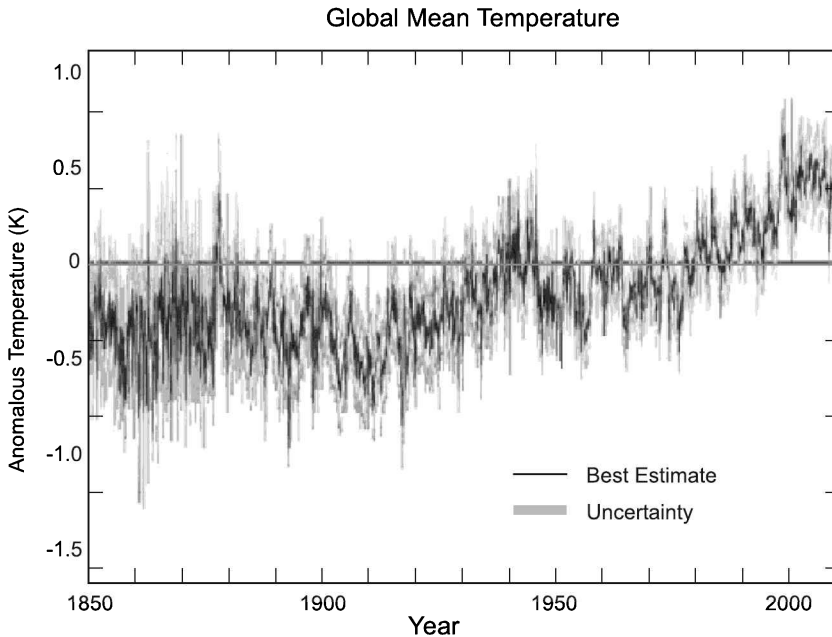


Figure 1.41 Anomalous Global Mean Temperature from surface stations (solid), relative to the mean over 1960–1990, along with estimates of uncertainty (shaded). *Source:* <http://hadobs.metoffice.com/hadcrut3>.

of temperature during the 1940s and 1950s coincides with a plateau in atmospheric CO_2 (Fig. 1.14). r_{CO_2} then actually decreased, even though anthropogenic emission accelerated during the period (Fig. 1.16). Then, in the late 1970s, GMT again increased steadily. That warming continued until the close of the twentieth century, after which GMT plateaued, declining slightly. This evolution is mirrored in the satellite record (Fig. 1.39).

Overall, the twentieth century was marked by three major episodes: 3 decades of steady warming, followed by 4 decades of steady cooling or no systematic change at all, followed by 2 decades of renewed steady warming. Together, those episodes led to net warming during the twentieth century of nearly 1.0 K. That deviation is twice as great as the one recorded during the latter half of the nineteenth century (albeit when measurements were sparse).

The systematic variation of GMT is modulated substantially by its climate variability. During the 1940s and 1950s, GMT decreased. Systematic warming over the twentieth century was then clearly overshadowed by climate variability. During the early twentieth century and late twentieth century, when GMT increased, it was, with equal likelihood, reinforced by climate variability. For this reason, the trend during an individual episode, even one lasting a couple of decades, represents not the secular variation or climate variability, but both. Only over a long enough interval for opposing swings of climate variability to cancel does a trend reflect the true secular variation of GMT. Between 1910 and 2010, GMT increased at a rate of ~ 0.09 K/decade. The mean trend over the twentieth century is ~ 0.075 K/decade (IPCC, 2007). These values are only about half of the warming trend that appeared during the last of the

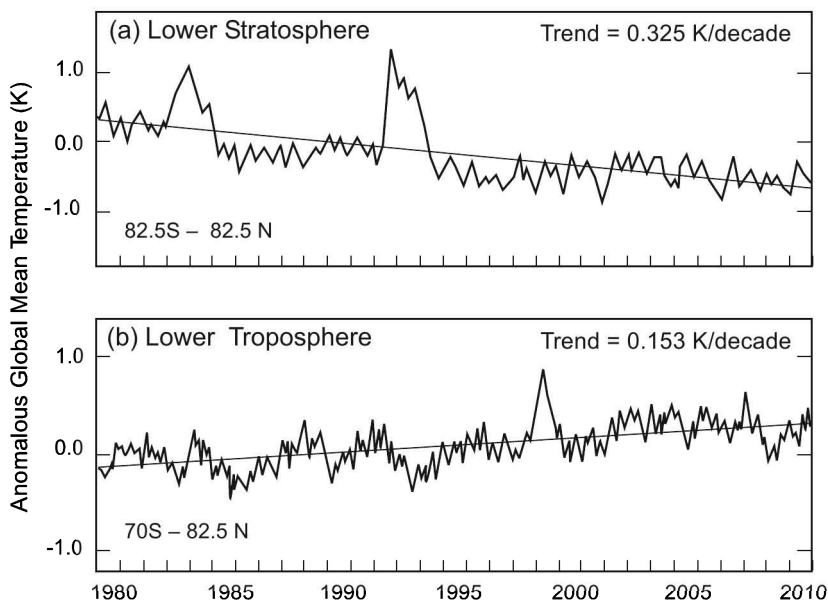


Figure 1.42 Satellite record of anomalous Global Mean Temperature in (a) lower stratosphere and (b) lower troposphere. Note that, in (b), the contribution from Antarctica is omitted. *Source:* www.ssmi.com/msu.

major episodes in Fig. 1.39 (e.g., during the 1980s and 1990s). It is about a third of the rate implied by Arrhenius' calculation, which did not account for adjustments of the climate system to increasing CO_2 .

Compared in Fig. 1.42 are the satellite records of GMT in the troposphere and stratosphere. In the lower troposphere (Fig. 1.42b), GMT is punctuated by warming during the 1997–1998 El Niño. A counterpart signature of the 1983 El Niño is less conspicuous. The tropospheric record exhibits a steady warming trend, of order 0.15 K/decade. (Noteworthy, however, is the exclusion of contributions from Antarctica, where the trend is negative; Fig. 1.40b,c.) In the lower stratosphere (Fig. 1.42a), even the 1997–1998 El Niño is absent. Instead, GMT in the stratosphere is punctuated by episodes of warming following the eruptions of El Chichón (1982) and Pinatubo (1991). Those episodes of warming are scarcely visible in the tropospheric record, which exhibits cooling following the eruptions. (cf. Fig. 1.27). In contrast to the warming trend in the troposphere, GMT in the stratosphere exhibits a cooling trend. Of order -0.30 K/decade, it is opposite to and about twice as strong as the warming trend in the lower troposphere.

1.6.2 Proxy records

Instrumental records before the nineteenth century are not available. In place of temperature measurements is a host of proxies, geological and other evidence that is used to infer temperature through a variety of techniques. These records are relied upon by students of *paleoclimate*, for periods before written records were

maintained. Such records are limited by coverage and uncertainties, which cloud their interpretation.

Figuring prominently among proxy records are ice cores (Fig. 1.12). Many come from Antarctica and Greenland, where deep cores extend far into the past. Through isotopic analysis, those cores provide estimates of prior temperature. Bubbles of air trapped inside the ice also provide estimates of atmospheric composition, including the concentrations of trace gases. Records like that in Fig. 1.12 indicate that, on long time scales, temperature and CO_2 vary dependently. They imply further that CO_2 lags temperature by 500–1000 years (Fischer et al., 1999; Indermuhle et al., 2000; Monnin et al., 2001). The lag suggests that, at least on millennial time scales, changes of temperature lead to changes of CO_2 (e.g., through changes in production at the Earth's surface and in exchanges with the ocean).

A close relationship between temperature and CO_2 is also evident on much shorter time scales, short enough to be resolved in the instrumental record. That relationship is hinted at by the plateau of r_{CO_2} during the 1940s and 1950s (Fig. 1.14), which coincided with a decline of GMT (Fig. 1.41). The budget of carbon is complex (Sec. 17.3). Poorly documented, individual sources and sinks may never be resolved in detail. Yet, their collective impact on atmospheric carbon is consolidated in a single property,

r_{CO_2} :
The rate of increase $\frac{d}{dt}r_{\text{CO}_2} = \dot{r}_{\text{CO}_2}$ must equal the net emission rate of CO_2 , collected from all sources and sinks. Plotted in Fig. 1.43a is the net emission rate of CO_2 (solid), evaluated from the record at Mauna Loa (Fig. 1.15). The behavior has been lowpass filtered to changes that occur on time scales longer than 2 years. By doing so, it discriminates to the hysteresis in the annual cycle of emission and absorption (Fig. 1.15) – the part that accumulates over successive years. Net emission is uniformly positive, consistent with CO_2 increasing monotonically. However, it varies substantially from one year to the next. About a mean of ~ 1.5 ppmv/yr, net emission changes by almost 100%: from twice as great in some years to nearly zero in others. These large changes reflect differences between years in how much CO_2 is emitted during one phase of its annual cycle versus how much is absorbed during the opposite phase. Of order 50–100%, the large change of net emission represents an imbalance between sources and sinks. By comparison, the contribution from fossil fuel emission varies only gradually (Rotty, 1987); cf. Fig. 1.16. Superimposed in Fig. 1.43 is the satellite record of global-mean temperature (Fig. 1.39), likewise lowpass filtered (dashed). Net emission of CO_2 closely tracks the evolution of GMT. Achieving a correlation of 0.80, the variation of GMT accounts for most of the variance in CO_2 emission.

Plotted in Fig. 1.43b is the rate of change in isotopic composition, $\frac{d}{dt}\delta^{13}\text{C} = \dot{\delta}^{13}\text{C}$ (solid).¹² Its mean is negative, consistent with the long-term decline of $\delta^{13}\text{C}$ in ice cores (Fig. 1.14). However, like emission of CO_2 , differential emission of $^{13}\text{CO}_2$ varies substantially from one year to the next. It too tracks the evolution of GMT – just out of phase. When GMT increases, emission of $^{13}\text{CO}_2$ decreases and vice versa. The records achieve a correlation of -0.86 . Hence the variation of GMT, which accounts for most of the variance in emission of CO_2 , also accounts for most of the variance in differential emission of $^{13}\text{CO}_2$.

¹² $\frac{d}{dt}\delta^{13}\text{C}$ is approximately equal to the percentage rate of change of $^{13}\text{CO}_2$ minus that of $^{12}\text{CO}_2$. It thus measures the rate of differential emission of $^{13}\text{CO}_2$.

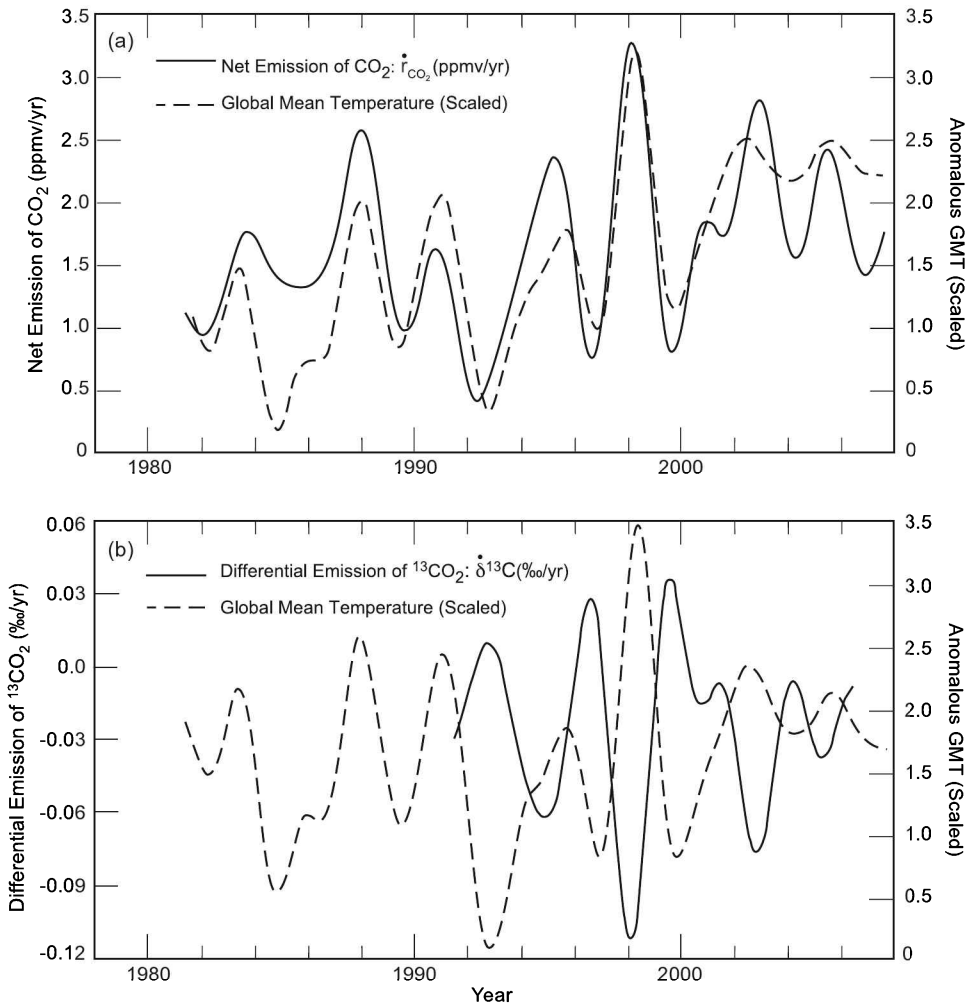


Figure 1.43 (a) Net emission rate of CO₂, $\dot{r}_{CO_2} = \frac{d}{dt} r_{CO_2}$ (ppmv/yr), derived from the Mauna Loa record (Fig. 1.15), lowpass filtered to changes that occur on time scales longer than 2 years (solid). Superimposed is the satellite record of anomalous Global Mean Temperature (Fig. 1.39), lowpass filtered likewise and scaled by 0.225 (dashed). Trend in GMT over 1979–2009 (not included) is ~ 0.125 K/decade. (b) As in (a), but for rate of change in isotopic carbon fraction, $\dot{\delta}^{13}C = \frac{d}{dt} \delta^{13}C$, likewise from Mauna Loa (cf. Fig. 1.14). $\frac{d}{dt} \delta^{13}C$ is approximately equal to the percentage rate of change of ¹³CO₂ minus that of ¹²CO₂. It thus measures the rate of differential emission of ¹³CO₂.

The out-of-phase relationship between r_{CO_2} and $\delta^{13}C$ in the instrumental record (Fig. 1.43) is the same one evidenced on longer time scales by ice cores (Fig. 1.14). The out-of-phase relationship in ice cores is regarded as a signature of anthropogenic emission, subject to uncertainties (Sec. 1.2.4). The out-of-phase relationship in the instrumental record, however, is clearly not anthropogenic. Swings of GMT following the eruption of Pinatubo and during the 1997–1998 El Niño were introduced through

natural mechanisms (cf. Figs 1.27; 17.19, 17.20). Changes in Fig. 1.43 reveal that net emission of CO₂, although ¹³C lean, is accelerated by increased surface temperature. Outgassing from ocean, which increases with temperature (Sec. 17.3), is consistent with the observed relationship – if the source region has anomalously low δ¹³C. So is the decomposition of organic matter derived from vegetation. Having δ¹³C comparable to that of fossil fuel, its decomposition is likewise accelerated by increased surface temperature.

Conspicuous in the records is the warming during the 1997–1998 El Niño, when GMT increased sharply. Intrinsic to the ocean, that perturbation to the climate system led to a surge in emission of CO₂ and a simultaneous reduction in differential emission of ¹³CO₂ – each by more than 100%. Such changes of global temperature cannot be explained by changes of carbon dioxide. Coherent with large swings in the net emission of CO₂ and ¹³CO₂, they imply that net transfer of carbon to the atmosphere involves processes that are sharply modulated by surface temperature. The respective sensitivity of those processes to surface temperature is established by coherent changes of net emission (Sec. 8.7.1). The dependence of CO₂ emission on temperature is poorly understood. It is not accounted for in GCMs, wherein atmospheric CO₂ is simply specified.

Older records of temperature and composition rest on proxy evidence, which is limited by coverage and uncertainty. Atmospheric properties inferred from ice cores are subject to a number of unknowns. At shallow depth, their layering resolves individual years. However, at greater depth, such features are blurred by diffusion, which limits temporal resolution. The isolation and stability of air trapped inside ice is a matter of ongoing debate.

So is the dating of features, which is pursued through half a dozen different techniques (e.g., Ruddiman and Raymo, 2003; Petit et al., 1999; Jouzel et al., 1996; Lorius et al., 1985; Sowers et al., 1993). Discrepancies in timing between those techniques range from centuries at shallow depth to millennia at greater depth. Those uncertainties are not allayed by cores from neighboring sites, which reveal similar discrepancies (Parrenin et al., 2007).

Dating of tree rings, *dendrochronology*, also figures prominently in reconstructions of previous climate. It relies on the width and number of annual growth cycles, in concert with radiocarbon dating of fossil specimens. The width of individual rings is used to infer temperature. Like ice cores, however, inferences drawn from those records are clouded by uncertainties; see Moberg et al. (2008) for an overview. In particular, temperature is but one of several environmental factors that influence annual tree growth (Bednarz and Ptak, 1990). Annual growth is also influenced by moisture, disease, and infestation. Another influence may be related more directly to its interpretation. Tree growth during the twentieth century differs distinctly from earlier growth (Briffa, 2000). It then diverges from the instrumental record – about the time that direct measurements of temperature became widespread. Why is not understood. Regardless of its origin, this feature seriously complicates the inference of temperature from annual tree growth. Supporting tree-ring evidence are sedimentary records from lakes and rivers. Because they form differently under frozen and unfrozen conditions, they too are clouded by uncertainties.

Changes in temperature inferred for previous climates are accompanied by changes in glaciation. Figure 1.12 compares, during the last 400,000 years, the record of temperature inferred from the Vostok core over Antarctica against global ice volume (reversed) inferred from sedimentary cores that were collected from around the globe. On time scales of millennia and longer, the two are strongly correlated. Gradual cooling

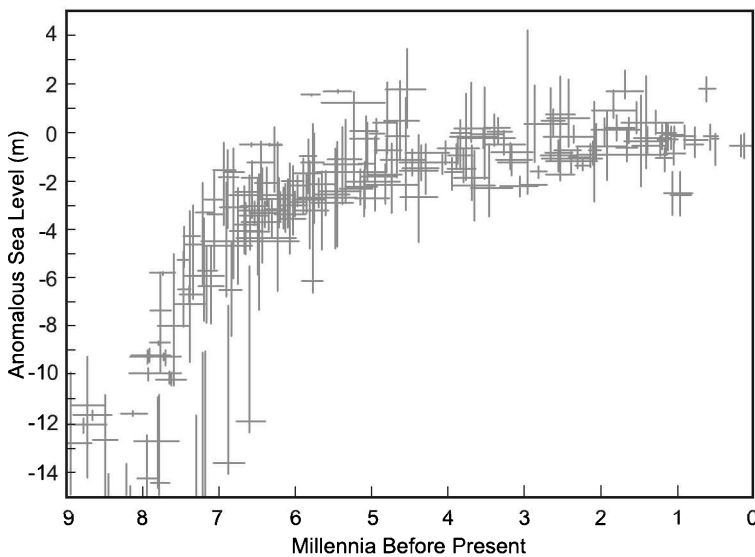


Figure 1.44 Reconstruction of sea level during the Holocene epoch, inferred from geological proxies at several sites around the Earth. The estimates rely on corrections for the vertical displacement of continents, which introduces one of several major uncertainties in dating (horizontal bar) and value (vertical bar). *Sources:* Fleming et al. (1998), Fleming (2000), and Milne et al. (2009).

toward temperature minima corresponds well with glacial periods. The comparatively abrupt warming to temperature maxima corresponds equally well with the approach to interglacial periods. Included is the current one, which defines the *Holocene* epoch.

The swings of ice volume in Fig. 1.12 imply reverse swings in sea level. To leading order, maritime ice does not contribute to such changes (Prob. 1.17). Sea level can then be influenced only by melting of continental ice. Plotted in Fig. 1.44 is a reconstruction of sea level during the last 10,000 years. It has been inferred from a variety of geological proxies, including sediments and coral skeletons that flourish at particular depths; Milne et al. (2009) provides a tabulation. Following the last glacial maximum 20,000 years ago, sea level rose sharply, at some 10 m/millennium. It then leveled off about 7000 years ago, although it continued to increase gradually. The attainment of those levels was contemporaneous with the development by humans of agriculture; the rise of major civilizations in the fertile crescent, Egypt, and Indus valley; and the onset of historical records. Since then, sea level appears to have risen by a couple of meters. However, the scatter of proxy evidence is noteworthy: about a meter.

Overall, the proxy record indicates a gradual increase of sea level during the last 5000 years, at ~ 4 cm/century. By comparison, sea level is currently rising at some 10–20 cm/century (Douglas, 1997). Most of this rise derives not from glacial melting, but from thermal expansion of the ocean (IPCC, 2007). The latter involves changes of temperature in the ocean interior, for which observations are scant (Chap. 17). It is noteworthy that other records during the same period as Fig. 1.44 reveal short-term fluctuations of sea level that are not resolved in such reconstructions (e.g., Siddall et al., 2003). As large as a couple of meters, those fluctuations operate on centennial time scales. A similar event some 14,000 years ago, the so-called *meltwater pulse*, involved a

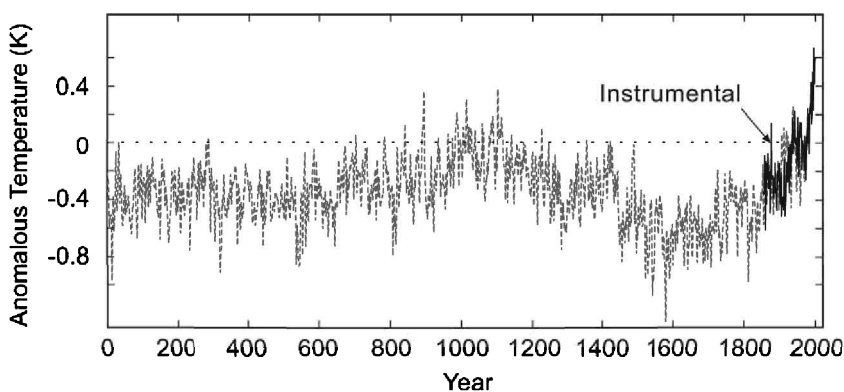


Figure 1.45 A reconstruction of Global Mean Temperature during last two millennia, inferred from a variety of proxies. After Moberg et al. (2005). Reprinted by permission of Macmillan Publishers Ltd: *Nature* (Moberg et al., 2005), Copyright (2005).

sea level rise of more than 10 m in just a couple of centuries (Webster, 2004; Stanford et al., 2006).

Presented in Fig. 1.45 is a reconstruction of temperature during the last two millennia based on a variety of proxies from the Northern Hemisphere, including tree rings. The record presented is but one of several such reconstructions, which are technique dependent (Moberg et al., 2008). Nonetheless, it illustrates major features that have long been recognized. After about the year 600, temperature warmed steadily, maximizing during the so-called *Medieval Warm Period* that prevailed during 1000–1300 (Lamb, 1965). The period is well documented in anecdotal evidence from the North Atlantic and Europe, in Norse explorations that were unhindered by ice, in the Viking colonization of Greenland, and in the northward advance of agriculture that enabled a wine industry to flourish in England (see, e.g., Lamb, 1982). The Medieval Warm Period is also evidenced by records of the tree line in the Alps, which ascended then by some 100 m.

Following the Medieval Warm Period was a temperature swing in the opposite direction. After the twelfth century, temperature decreased steadily – for several centuries. Minimum values were attained during the so-called *Little Ice Age*, which prevailed between the sixteenth and nineteenth centuries. The Little Ice Age is chronicled in anecdotal evidence from Europe and North America, surrounding the southward advance of pack ice into the North Atlantic; the expansion of glaciers, which descended some 100 m lower than their extent in the late twentieth century; and the freezing of major bodies of water like the Thames, New York harbor, and parts of the Baltic and Bosphorus. Then, in the nineteenth century; temperature began increasing. It eventually merged with the increase that appears in the instrumental record, which commenced later that century. How much temperature rebounded, like so many features in proxy evidence, depends on which reconstruction is adopted.

Whether the Medieval Warm Period and Little Ice Age were global events or even existed has been questioned. Uncertainties over their existence surround the veracity and interpretation of tree-ring records. However, reconstructions exclusive of tree-ring evidence produce much the same pattern as in Fig. 1.45, which is a robust feature

of the proxy record (Loehle, 2007; Loehle and McCulloch, 2008). The Medieval Warm Period and Little Ice Age have also been documented in records worldwide, including moraine deposits, sedimentary records, and ice cores (e.g., Wilson et al., 1979; Grove and Switsur, 1994; Broecker, 2000; Kreutz et al., 1997; Adhikari and Kumon, 2001; Cook et al., 2002; Hall and Denton, 2002; McGann, 2008; Williams et al., 2007; Araneda et al., 2007). Those records evidence a global retreat of glaciers during the Medieval Warm Period, followed by their systematic advance during the approach to the Little Ice Age.

How much temperature was elevated during the Medieval Warm Period, like many features of temperature reconstructions, is a matter of debate. The reconstruction in Fig. 1.45 implies that temperature then was about 0.5 K warmer than previously. Other reconstructions place temperature then 1.0–2.0 K warmer (Lamb, 1965; Keigwin, 1996; Huang and Pollack, 1997; Loehle, 2007). That would make the Medieval Warm Period as warm or warmer than temperature during the late twentieth century.

Similar uncertainty surrounds how fast temperature rebounded following the Little Ice Age. Because, on long time scales, CO_2 is coupled to temperature, the same uncertainty applies to its increase. The reconstruction in Fig. 1.45 implies warming over the last two centuries of 0.03–0.05 K/decade (e.g., Moberg et al., 2005). Reconstructions free of tree-ring uncertainty imply similar warming since the Little Ice Age (Loehle, 2007). Representing the average trend over two centuries, those estimates of natural warming are about half of the warming trend that appears in the instrumental record during the twentieth century.

SUGGESTED REFERENCES

The Boltzmann distribution is developed along with other components of molecular kinetics in *Statistical Thermodynamics* (1973) by Lee, Sears, and Turcotte.

Aeronomy of the Middle Atmosphere (1986) by Brasseur and Solomon includes a comprehensive treatment of anthropogenic trace gases.

The Intergovernmental Panel Report on Climate Change (2007) provides an overview of the climate problem, numerical modeling of it, and historical evidence.

A comprehensive treatment of the global energy budget and its relationship to the general circulation is presented in *Global Physical Climatology* (1994) by Hartmann.

PROBLEMS

1. Derive expression (1.15) for the molar fraction of the i th species in terms of its mass mixing ratio.
2. Derive an expression for the volume mixing ratio of the i th species in terms of its mass mixing ratio.
3. Relate the volume mixing ratio of the i th species to the fractional abundance of i molecules present.
4. Show that the sum of partial volumes equals the total volume occupied by a mixture of gases (1.6).
5. Demonstrate that 1 atm of pressure is equivalent to that exerted by (a) a 760 mm column of mercury, (b) a 32' column of water. (The density of Hg at 273 K is $1.36 \cdot 10^4 \text{ kg m}^{-3}$.)
6. Consider moist air at a pressure of 1 atm and a temperature of 20°C. Under these conditions, a relative humidity of 100% corresponds to a vapor pressure

of 23.4 hPa. At that vapor pressure, air holds the maximum abundance of water vapor possible under the foregoing conditions and is said to be *saturated*. For this mixture, determine (a) the molar fraction N_{H_2O} , (b) the mean molar weight \bar{M} , (c) the mean specific gas constant \bar{R} , (d) the absolute concentration of vapor ρ_{H_2O} , (e) the mass mixing ratio of vapor r_{H_2O} , (f) the volume mixing ratio of vapor.

7. From the distribution of water vapor mixing ratio shown in Fig. 1.17, plot the vertical profile of water vapor number density $[H_2O](z)$ at (a) the equator, (b) 60°N.
8. In terms of the vertical profile of water vapor mixing ratio $r_{H_2O}(z)$, derive an expression for the total precipitable water vapor Σ_{H_2O} in mm.
9. If

$$r_{H_2O}(\phi, z) = r_0(\phi)e^{-\frac{z}{h}},$$

where $h = 3$ km and $r_0(\phi)$ is the zonal-mean mixing ratio at the surface and latitude ϕ in Fig. 1.17, use the results of Prob. 1.8 to compute the total precipitable water vapor at (a) the equator, (b) 60°N.

10. From the mixing ratio of ozone r_{O_3} shown in Fig. 1.20, plot the vertical profile of ozone number density $[O_3](z)$ above 100 hPa at (a) the equator, (b) 60°N.
11. As in Prob. 1.10a, but plot the contribution to total ozone $\Sigma_{O_3}(z)$ from 100 hPa to an altitude z over the equator.
12. Cloud contributes much of the Earth's albedo, which has a value of about 0.30. Calculate the change of equivalent blackbody temperature corresponding to a 5% reduction of albedo.
13. The Earth's equivalent blackbody temperature, $T_e = 255$ K, corresponds to a level in the middle troposphere, above the layer in which water vapor is concentrated. This also corresponds to the level from which most of the outgoing LW radiation is emitted to space. Suppose atmospheric water vapor increases to raise the top of the (optically thick) water vapor layer by 1 km, while the albedo remains unchanged. (a) What is the corresponding change of the Earth's equivalent blackbody temperature? (b) If a lapse rate of 6.5 K km^{-1} is maintained inside the water vapor layer, by how much will the Earth's mean surface temperature change?
14. By appealing to the hydrostatic relationship (1.17), contrast the vertical spacing of isobaric surfaces ($p = \text{const}$) at a horizontal site where air is cold from one where air is warm.
15. The mean pressure observed in a planetary atmosphere varies with height as

$$p(z) = \frac{p_0}{1 + \left(\frac{z}{h}\right)^2}.$$

Determine the mean variation of temperature with height.

16. Venus has a mean distance from the sun of 0.7 that of Earth and an albedo of 0.75. Calculate the radiative-equilibrium temperature of Venus. Compare the result against the observed surface temperature of 750 K.
17. (a) Apply Archimedes Principle to argue that maritime ice, if it melts, does not alter sea level. (b) The density of water is actually greatest at 4°C (1000 kg/m³), some 8% greater than at 0°C (916.8 kg/m³). If maritime ice occupies 10% of the surface area of ocean and if the ocean has a mean depth of 2 km, estimate the mean rise

of sea level that would result if all maritime ice underwent a transformation of state from 0°C to the temperature 4°C of surrounding sea water. (c) Maritime ice is almost pure water, which, at 4°C , has a density some 2.5% less than sea water (1027 kg/m^3). Under the same circumstances, repeat the calculation in (b).

18. Lower-tropospheric temperature measured from space by the microwave instruments MSU and AMSU represents, not temperature at the earth's surface, but temperature that is averaged vertically over the lowest $\sim 5\text{ km}$ of atmosphere. As horizontal structure is smoother at upper levels than at the surface, vertical averaging blurs features that are present in the horizontal distribution of surface temperature. Suppose that the blurring of horizontal structure is described by diffusion-like equation

$$\frac{\partial T}{\partial z} = k(z)\nabla^2 T,$$

where $T = T(z)$ is the temperature at altitude z , ∇ denotes horizontal gradient, and $k(z)$ is the effective diffusivity. According to this relation, T changes with altitude in proportion to the horizontal curvature of T , which is thereby damped with increasing altitude. Show that, although horizontal structure becomes increasingly blurred with altitude, its global mean does not change.

Thermodynamics of gases

The link between the circulation and transfers of energy from the Earth's surface is thermodynamics. Thermodynamics deals with internal transformations of the energy of a system and exchanges of energy between that system and its environment. Here, we develop the principles of thermodynamics for a discrete system, namely an air parcel moving through the circulation. In Chap. 10, these principles are generalized to a continuum of such systems, which represents the atmosphere as a whole.

2.1 THERMODYNAMIC CONCEPTS

A *thermodynamic system* refers to a specified collection of matter (Fig. 2.1). Such a system is said to be “closed” if no mass is exchanged with its surroundings. Otherwise it is “open.” The air parcel that will serve as our system is, in principle, closed. In practice, however, mass can be exchanged with the surroundings through entrainment and mixing across the system's boundary, which is referred to as the *control surface*. In addition, trace constituents such as water vapor can be absorbed through diffusion across the control surface. Above the planetary boundary layer, such exchanges are slow compared with other processes that influence an air parcel. The system may therefore be treated as closed.

The thermodynamic state of a system is defined by the various properties characterizing it. In a strict sense, all of those properties must be specified to define the system's thermodynamic state. However, that requirement is simplified for many applications, as is discussed next.

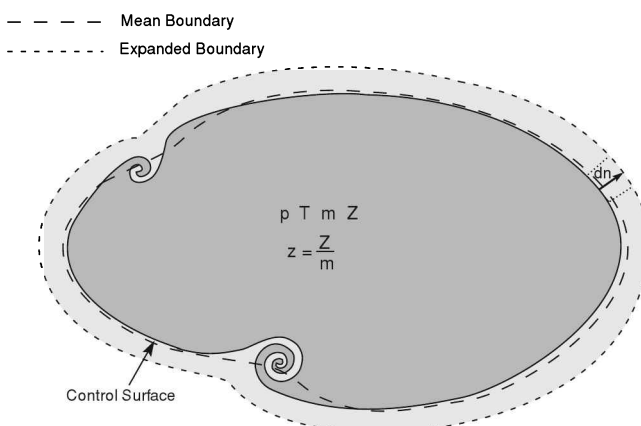


Figure 2.1 A specified collection of matter defining an infinitesimal air parcel. The system's thermodynamic state is characterized by the extensive properties m and Z and by the intensive properties p , T , and z .

2.1.1 Thermodynamic properties

Two types of properties characterize the state of a system. A property that does not depend on the mass of the system is said to be *intensive*. Otherwise it is *extensive*. Intensive and extensive properties are usually denoted with lower and upper case symbols, respectively. Pressure and temperature are examples of intensive properties. Volume is an extensive property. An intensive property z may be defined from an extensive property Z , by referencing the latter to the mass m of the system

$$z = \frac{Z}{m}. \quad (2.1)$$

The intensive property is then referred to as a *specific property*. The specific volume $v = \frac{V}{m}$ is an example. If a system's properties do not vary in space, it is said to be *homogeneous*. Otherwise it is *heterogeneous*. Because an air parcel is of infinitesimal dimension, it is by definition homogeneous (so long as it involves only gas phase). On the other hand, stratification of density and pressure make the atmosphere as a whole a heterogeneous system.

A system can exchange energy with its surroundings through two fundamental mechanisms. It can perform work on its surroundings, which represents a mechanical exchange of energy with the environment. In addition, heat can be transferred across the control surface, which represents a thermal exchange of energy with the environment.

2.1.2 Expansion work

The system relevant to the atmosphere is a compressible gas, perhaps containing an aerosol of liquid and solid particles. For this reason, the primary mechanical means of exchanging energy with the environment is expansion work (see Fig. 2.1). If a pressure imbalance exists across the control surface, the system is out of mechanical equilibrium. It will automatically expand or contract to relieve the mechanical imbalance. By

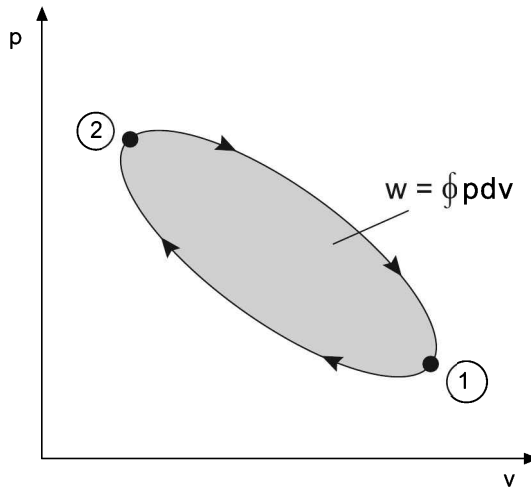


Figure 2.2 Expansion work performed by the system during a cyclic process, wherein it is restored to its initial state.

adjusting to the environmental pressure, the system performs expansion work or has such work performed on it. The incremental expansion work δW performed by displacing a section of control surface $dS = dSn$, with unit normal \mathbf{n} , perpendicular to itself by $d\mathbf{n}$ is

$$\begin{aligned}\delta W &= (p d\mathbf{S}) \cdot d\mathbf{n} \\ &= p dV,\end{aligned}\tag{2.2}$$

where $dV = dS dn$ is the incremental volume displaced by the section dS . Integrating (2.2) between the initial and final volumes yields the net work performed “by” the system

$$W_{12} = \int_{V_1}^{V_2} p dV.\tag{2.3}$$

Dividing both sides by the mass of the system then gives the specific work

$$w_{12} = \int_{v_1}^{v_2} p dv.\tag{2.4.1}$$

For a cyclic variation, namely one in which the initial and final states of the system are identical,

$$w = \oint p dv.\tag{2.4.2}$$

The cyclic work (2.4.2) is represented in the area enclosed by the path of the system in the $p - v$ plane (Fig. 2.2). As the system is restored to its initial state, the net change of properties is zero. However, the same is not true of the work performed during the cycle. Unless the system returns to its initial state along the same path it followed out of that state, net work is performed during a cyclic variation.

On a fluid system, another form of work can be performed, one analogous to stirring. *Paddle work* corresponds to a rearrangement of fluid, as would occur by turning a paddle immersed in the system. The reaction of the fluid against the paddle must be overcome by a torque, which in turn performs work when exerted across an angular displacement. Although secondary to expansion work, paddle work has applications to dissipative processes associated with turbulence (cf. Fig. 2.1).

2.1.3 Heat transfer

Energy can also be exchanged thermally, via heat transfer Q across the system's control surface. If an air parcel moves into an environment of a higher temperature, it will absorb heat from its surroundings (e.g., through diffusion or thermal conduction). If the system is open, a similar process can occur through the absorption of water vapor from the surroundings, followed by condensation and the release of latent heat. Heat can also be exchanged through radiative transfer. If it interacts radiatively with surroundings at a higher temperature, an air parcel will absorb more radiant energy than it emits.

For many applications, heat transfer is secondary to processes that are introduced through motion, which operates on time scales of a day and shorter. This is especially true above the boundary layer, where turbulent mixing between bodies of air is relatively weak. In the free atmosphere and away from cloud, the prevailing form of heat transfer is radiative. Operating on a time scale of order 2 weeks, radiative transfer is slow by comparison with expansion work, which influences an air parcel on time scales of only a day.

If no heat is exchanged between a system and its environment, the control surface is said to be *adiabatic*. Otherwise, it is *diabatic*. Because, for many applications, heat transfer is slow compared with other processes that influence a parcel, adiabatic behavior is a good approximation. Obviously, heat transfer must be central to processes that operate on long time scales, as it is ultimately responsible for driving the atmosphere into motion. For this reason, diabatic effects prove to be important for the long-term maintenance of the general circulation, even though they can be ignored for behavior that operates on shorter time scales.

2.1.4 State variables and thermodynamic processes

In general, describing the thermodynamic state of a system requires specifying all of its properties. However, that requirement can be relaxed for gases and other substances at normal temperatures and pressures. A *pure substance* is one whose thermodynamic state is uniquely determined by any two intensive properties. For such a system, intensive properties are called *state variables*. From any two state variables z_1 and z_2 , a third z_3 may be determined through an *equation of state* of the form

$$f(z_1, z_2, z_3) = 0,$$

or, equivalently,

$$z_3 = g(z_1, z_2).$$

In this sense, a pure substance has two *thermodynamic degrees of freedom*. Any two state variables uniquely specify its thermodynamic state. An ideal gas is a pure substance. It has as its equation of state the ideal gas law (1.1). Because no more than two intensive properties are independent, the state of a pure substance may be specified in the plane of any two of its state variables (z_1, z_2). The latter represents the *state space* of the system. It defines the collection of all possible thermodynamic states. For instance, the state space of an ideal gas may be represented by the $v - T$ plane. Any third state variable is then described by a surface, as a function of the two independent state variables. For example, $p = p(v, T)$, as illustrated in Fig. 2.3 for an ideal gas.

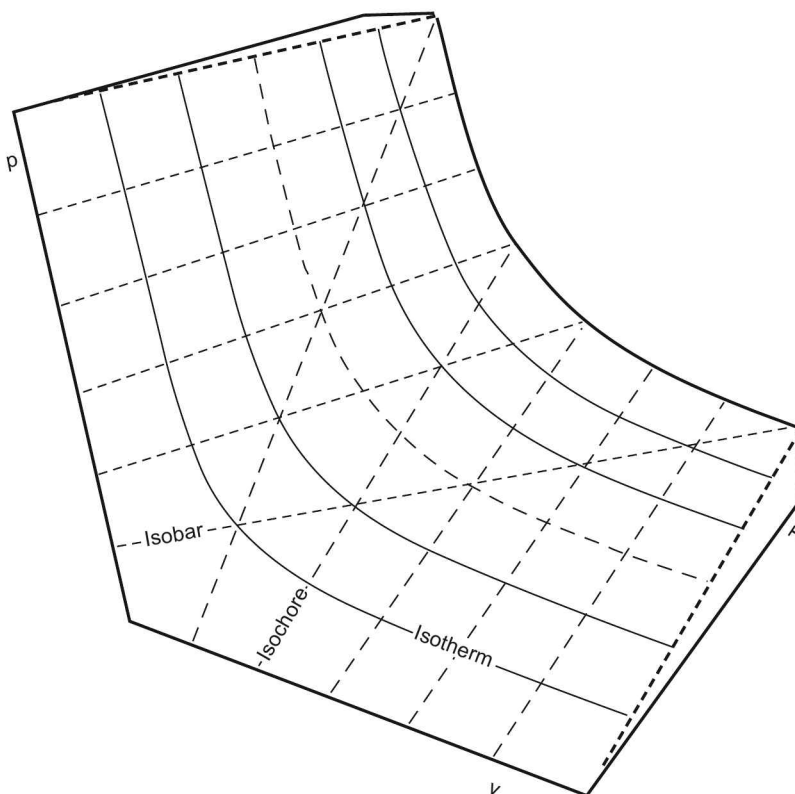


Figure 2.3 State space of an ideal gas, illustrated in terms of p as a function of the state variables v and T . Superposed are contours of constant pressure (*isobars*), constant temperature (*isotherms*), and constant volume (*isochores*).

A homogeneous system is said to be in *mechanical equilibrium* if there exists at most an infinitesimal pressure difference between the system and its surroundings (unless its control surface happens to be rigid). Likewise, a homogeneous system is said to be in *thermal equilibrium* if there exists at most an infinitesimal temperature difference between the system and its surroundings (unless its control surface happens to be adiabatic). If it is in mechanical equilibrium and in thermal equilibrium, a homogeneous system is said to be in *thermodynamic equilibrium*.

Thermodynamics addresses how a system evolves from one state to another (e.g., from one position in state space to another position). The transformation of a system between two states is referred to as a *thermodynamic process*. It describes a path in state space. As depicted in Fig. 2.4, two thermodynamic states are connected by infinitely many paths. Consequently, defining a process requires specifying the particular path in state space along which the system evolves. In contrast, the change of a state variable depends on only the initial and final states of the system. It is therefore “path independent.” As illustrated in Fig. 2.2, a cyclic process may be decomposed into two legs: a forward leg, from the initial state to an intermediate state, and a reverse leg, from that intermediate state back to the initial state. Because it is path independent, the change of a state variable z along the reverse leg must equal minus the change

along the forward leg. Therefore, the cyclic integral of a state variable vanishes

$$\oint dz = 0. \quad (2.5)$$

The incremental change of a quantity $z(x, y)$ that satisfies (2.5) represents an exact differential

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy, \quad (2.6)$$

a description which holds only under certain conditions.

Exact Differential Theorem: Consider two continuously differentiable functions $M(x, y)$ and $N(x, y)$ in a simply-connected region of the $x - y$ plane. The contour integral between two points (x_0, y_0) and (x, y)

$$\int_{(x_0, y_0)}^{(x, y)} M(x', y') dx' + N(x', y') dy' \quad (2.7)$$

(e.g., along a specified curve $g(x, y)$ that relates x and y) is path independent if and only if

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}. \quad (2.8)$$

Under these circumstances, the quantity

$$M(x, y)dx + N(x, y)dy = dz$$

represents an exact differential. That is, there exists a function z such that

$$\begin{aligned} M(x, y) &= \frac{\partial z}{\partial x} \\ N(x, y) &= \frac{\partial z}{\partial y}. \end{aligned} \quad (2.9.1)$$

The contour integral (2.7) then reduces to

$$\begin{aligned} \int_{z(x_0, y_0)}^{z(x, y)} dz &= z(x, y) - z(x_0, y_0) \\ &= \Delta z \end{aligned} \quad (2.9.2)$$

or

$$z = z(x, y) + c. \quad (2.9.3)$$

The variable defined by (2.9) is a *point function*: It depends only on the evaluation point (x, y) . Similarly, its net change along a contour depends only on the initial and final points of the contour. Because it is unique only up to an additive constant, only the change of a point function is significant. Also referred to as a *potential function*, $z(x, y)$ defines an *irrotational* vector field

$$\begin{aligned} \mathbf{v} &= \nabla z \\ &= M(x, y)\mathbf{i} + N(x, y)\mathbf{j}, \end{aligned} \quad (2.10.1)$$

which satisfies

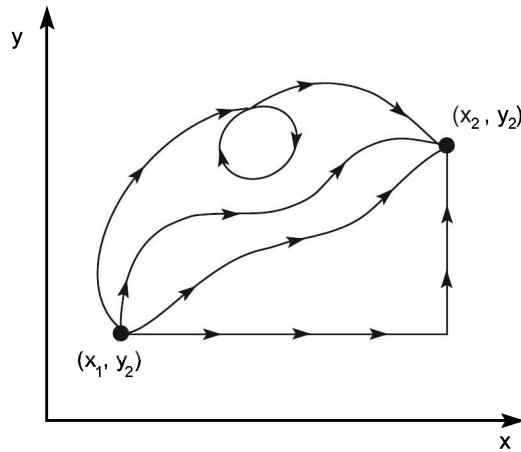


Figure 2.4 Possible thermodynamic processes between two states: (x_1, y_1) and (x_2, y_2) .

$$\nabla \times \mathbf{v} \equiv 0.^1 \quad (2.10.2)$$

An irrotational vector field may be represented as the gradient of a scalar potential z . Gravity \mathbf{g} is an irrotational vector field: $\nabla \times \mathbf{g} \equiv 0$. The work performed to displace a unit mass between two points in a gravitational field is independent of path. It defines the gravitational potential Φ .

Thermodynamic state variables are point functions. Properties of the system, they depend only on the system's state, not on its history. By contrast, the work performed by the system and the heat transferred into it during a thermodynamic process are not properties of the system. Work and heat transfer are, in general, *path functions*. They depend on the path in state space that is followed by the system. For this reason, the thermodynamic process must be specified to unambiguously define those quantities.

Path dependent, work and heat transfer can differ along the forward and reverse legs of a cyclic process. Consequently, the net work and heat transfer during a cyclic variation of the system need not vanish, as does the net change of a state variable (2.5). The path dependence of work and heat transfer produce a *hysteresis* in the $w - q$ plane, illustrated in Fig. 2.5. During a cyclic variation of the system, the cumulative work and heat

$$w = \int \delta w$$

$$q = \int \delta q$$

do not return to their original positions after the system has been restored to its initial state. The discrepancy $\Delta w = \oint \delta w$ after completing a cycle equals the area in the $p - v$ plane enclosed by the contour in Fig. 2.2. Due to hysteresis, successive cycles lead to a drift of the above quantities in the $w - q$ plane. The drift reflects the net work performed by the system and the net heat transferred into it.

Under special circumstances, the work performed by a system or the heat transferred into it "are" independent of path. Work and heat transfer then vanish for a

¹ If it is a force, the vector field \mathbf{v} is said to be *conservative* because the net work performed along a cyclic path vanishes (2.5).

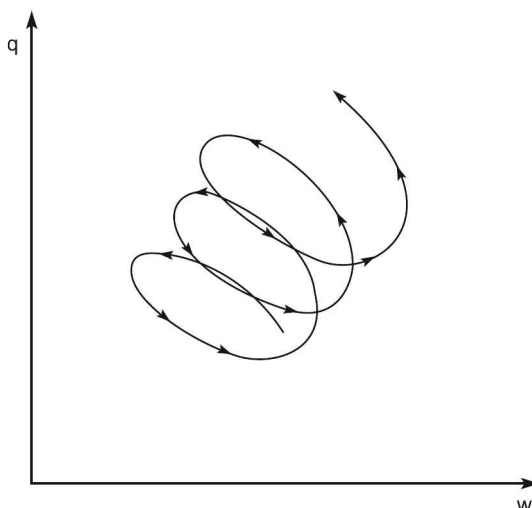


Figure 2.5 Cumulative work and heat transfer during successive thermodynamic cycles of the system. The path dependence of w and q introduces a hysteresis into those quantities, even though the system (i.e., the collection of state variables) is restored to its initial state after each cycle.

cyclic process. Each is therefore a point function. This special form of work or heat transfer may be used to define a state variable. For instance, the displacement work in a gravitational field is path independent. That work can therefore be used to define the gravitational potential Φ , which is a state variable (Sec. 6.2).

2.2 THE FIRST LAW

2.2.1 Internal energy

The First Law of Thermodynamics is inspired by an empirical finding: The work performed on an adiabatic system is independent of the process. That is, under adiabatic conditions, the work is independent of the path in state space followed by the system. For an adiabatic process, expansion work depends only on the initial and final states of the system. It therefore behaves as a state variable. The *internal energy* u is defined as that state variable whose change, under adiabatic conditions, equals the work performed on the system or minus the work performed “by” the system.

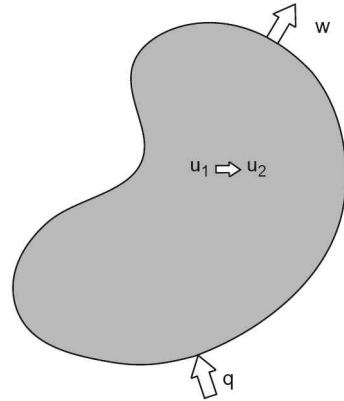
$$\Delta u = -w_{ad}. \quad (2.11.1)$$

For an incremental process,

$$du = -\delta w_{ad}, \quad (2.11.2)$$

where δ describes an incremental change that is, in general, “path dependent” and d to one that is “path independent” (i.e., of a state variable). Under the special circumstances defined in the preceding paragraph, work is path independent. The net work performed

Figure 2.6 An air parcel undergoing a process between states 1 and 2, during which it performs work w and absorbs heat q .



during a cyclic process therefore vanishes. The system must then return to its initial state along the same path that it followed out of that state.

2.2.2 Diabatic changes of state

Consider now a diabatic process, as illustrated schematically in Fig. 2.6. Because heat is exchanged with the environment,

$$w \neq w_{ad} = -\Delta u.$$

The work performed by the system w will differ from that performed under adiabatic conditions, w_{ad} . The difference, $w - w_{ad} = q$, equals the energy transferred “into” the system through heat exchange (e.g., it measures the departure from adiabatic conditions). The change of internal energy is then described by

$$\Delta u = q - w. \quad (2.12.1)$$

Equation (2.12.1) constitutes the *First Law of Thermodynamics*. For an incremental process and for a system capable of expansion work only, the First Law may be expressed

$$du = \delta q - pdv. \quad (2.12.2)$$

In (2.12), pdv represents the work performed “by” the system on its surroundings and q the heat transfer “into” the system. The change of internal energy between two states is path independent. However, the same is not true of the work performed by the system and the heat transferred into it during the change of state. Except under adiabatic conditions, the work performed by the system during a change of state depends on the thermodynamic process, namely, the path in state space along which the system evolves. The same applies to the heat transferred into the system during the change of state. Because each depends on path, neither the work nor the heat transfer vanishes for a cyclic process. However, the change of internal energy does vanish for a cyclic process. The First Law therefore reduces to

$$\oint pdv = \oint \delta q. \quad (2.13)$$

During a cyclic process, the net work performed by the system (energy out) is balanced by the net heat absorbed by the system (energy in).

A closed system that performs work through a conversion of heat that is absorbed by it is a *heat engine*. Conversely, a system that rejects heat through a conversion of work that is performed on it is a *refrigerator*. In Chap. 6, we will see that individual air parcels comprising the circulation of the troposphere behave as a heat engine. By absorbing heat at the Earth's surface, through transfers of radiative, sensible, and latent heat, individual parcels perform net work as they evolve through a thermodynamic cycle (2.13). Ultimately realized as kinetic energy, the heat absorbed maintains the circulation against frictional dissipation. It makes the circulation of the troposphere thermally driven.

In contrast, the circulation of the stratosphere behaves as a radiative refrigerator. For motion to occur, individual air parcels must have work performed on them. The kinetic energy produced is eventually converted to heat and rejected to space through LW cooling. It makes the circulation of the stratosphere mechanically driven. Gravity waves and planetary waves that propagate upward from the troposphere are dissipated in the stratosphere. Their absorption exerts an influence on the stratosphere analogous to paddle work. By forcing motion that rearranges air, it drives the stratospheric circulation out of radiative equilibrium, which results in net LW cooling to space.

It is convenient to introduce the state variable *enthalpy*

$$h = u + pv. \quad (2.14)$$

In terms of enthalpy, the First Law becomes

$$dh = \delta q + v dp. \quad (2.15)$$

Enthalpy is useful for diagnosing processes that occur at constant pressure. Under those circumstances, the First Law reduces to a statement that the change in enthalpy equals the heat transferred into the system.

2.3 HEAT CAPACITY

Observations indicate that the heat absorbed by a homogeneous system which is maintained at constant pressure or at constant volume is proportional to the change of the system's temperature. The constants of proportionality between heat absorption and temperature change define the *specific heat capacity at constant pressure*

$$c_p = \frac{\delta q_p}{dT} \quad (2.16.1)$$

and the *specific heat capacity at constant volume*

$$c_v = \frac{\delta q_v}{dT}, \quad (2.16.2)$$

where the subscripts denote processes that are *isobaric* ($p = \text{const}$) and *isochoric* ($v = \text{const}$).

The specific heat capacities are related closely to the internal energy and enthalpy of the system. Because they are state variables, u and h may be expressed in terms of any two other state variables. For instance, $u = u(v, T)$, in which case

$$du = \left(\frac{\partial u}{\partial v} \right)_T dv + \left(\frac{\partial u}{\partial T} \right)_v dT$$

Incorporating this transforms the First Law (2.12.2) into

$$\left(\frac{\partial u}{\partial T}\right)_v dT + \left[\left(\frac{\partial u}{\partial v}\right)_T + p\right] dv = \delta q.$$

For an isochoric process, this reduces to

$$\begin{aligned} \left(\frac{\partial u}{\partial T}\right)_v &= \frac{\delta q_v}{dT} \\ &= c_v. \end{aligned} \quad (2.17.1)$$

Thus, the specific heat capacity at constant volume measures the rate at which internal energy increases with temperature during an isochoric process. In similar fashion, the enthalpy may be expressed $h = h(p, T)$, with which the First Law (2.15) becomes

$$\left[\left(\frac{\partial h}{\partial T}\right)_T - v\right] dp + \left(\frac{\partial h}{\partial T}\right)_p dT = \delta q.$$

For an isobaric process, this reduces to

$$\begin{aligned} \left(\frac{\partial h}{\partial T}\right)_p &= \frac{\delta q_p}{dT} \\ &= c_p. \end{aligned} \quad (2.17.2)$$

The specific heat capacity at constant pressure measures the rate at which enthalpy increases with temperature during an isobaric process.

In a strict sense, c_p and c_v are state variables. Hence they too depend on pressure and temperature. However, over ranges of pressure and temperature relevant to the atmosphere, the specific heats may be regarded as constant. Therefore, the change of internal energy during an isochoric process is proportional to the change of temperature alone (c_v) and similarly for the change of enthalpy during an isobaric process (c_p). For an ideal gas, the relationships (2.17) turn out to hold irrespective of process.

Consider the internal energy of an ideal gas in the form $u = u(p, T)$. Joule's experiment, which is pictured in Fig. 2.7, demonstrates that u is then a function of temperature alone. Two ideal gases that are initially isolated and at pressures p_1 and p_2 are brought into contact and allowed to equilibrate (e.g., by rupturing a diaphragm that separates them). Observations indicate that, during this process, no heat transfer takes place with the environment. The First Law then reduces to a statement that the change of internal energy equals minus the work performed. However, the volume of the system (that occupied collectively by both gases) does not change. Consequently, the work also vanishes, leaving $\Delta u = 0$. Yet, the final equilibrated pressure clearly differs from the initial pressures of the individual gases in isolation. Because the internal energy of the system does not change, it follows that u is not a function of pressure. Hence, for an ideal gas, $u(p, T)$ reduces to

$$u = u(T).$$

As it involves only state variables, this relationship must hold irrespective of process. Incorporating it and the ideal gas law into (2.14) demonstrates that the enthalpy is likewise a function of temperature alone

$$h = h(T).$$

Involving only state variables, it too holds irrespective of process.

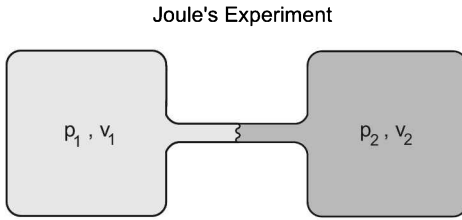


Figure 2.7 Schematic of Joule's experiment: Two ideal gases in states 1 and 2 are brought into contact by rupturing a diaphragm that separates them initially. No heat transfer with the surroundings is observed.

These conclusions imply that, for an ideal gas, equations (2.17) are valid in general. Integrating with respect to temperature yields finite values of internal energy and enthalpy, which are unique up to constants of integration. It is customary to define u and h so that they vanish at a temperature of absolute zero. With that convention,

$$u = c_v T \quad (2.18.1)$$

$$h = c_p T. \quad (2.18.2)$$

It follows that

$$c_p - c_v = R. \quad (2.19)$$

According to statistical mechanics, the specific heat at constant volume is given by

$$c_v = \frac{3}{2}R \quad (2.20.1)$$

for a monatomic gas, and by

$$c_v = \frac{5}{2}R \quad (2.20.2)$$

for a diatomic gas (e.g., Lee, Sears, and Turcotte, 1973). These values are confirmed experimentally over a wide range of pressure and temperature relevant to the atmosphere. Taking air to be chiefly diatomic, together with the value of R_d in (1.18), yields the specific heats for dry air

$$\begin{aligned} c_{vd} &= 717.5 \text{ Jkg}^{-1}\text{K}^{-1} \\ c_{pd} &= 1004.5 \text{ Jkg}^{-1}\text{K}^{-1} \end{aligned} \quad (2.21.1)$$

and the dimensionless constants

$$\gamma = \frac{c_p}{c_v} = 1.4 \quad (2.21.2)$$

$$\begin{aligned} \kappa &= \frac{R}{c_p} \\ &= \frac{\gamma - 1}{\gamma} \\ &\cong 0.286. \end{aligned} \quad (2.21.3)$$

With the aforementioned definitions, the First Law may be expressed in the two equivalent forms

$$c_v dT + p dv = \delta q \quad (2.22.1)$$

$$c_p dT - v dp = \delta q. \quad (2.22.2)$$

For isochoric and isobaric processes, these expressions reduce to

$$\delta q_v = c_v dT \quad (2.23.1)$$

$$\delta q_p = c_p dT, \quad (2.23.2)$$

respectively. Because the right-hand sides of (2.23) involve only state variables, the same must be true of the left-hand sides. Hence, under these special circumstances, heat transfer behaves as a state variable. Although generally path dependent, heat transfer during an isochoric process or during an isobaric process is uniquely determined by the change of temperature.

2.4 ADIABATIC PROCESSES

For an adiabatic process, the First Law reduces to

$$c_v dT + p dv = 0 \quad (2.24.1)$$

$$c_p dT - v dp = 0. \quad (2.24.2)$$

Dividing through by T and introducing the gas law transforms (2.24) into

$$c_v d \ln T + R d \ln v = 0 \quad (2.25.1)$$

$$c_p d \ln T - R d \ln p = 0. \quad (2.25.2)$$

Equations (2.25) may be integrated to obtain the identities

$$T^{c_v} v^R = \text{const} \quad (2.26.1)$$

$$T^{c_p} p^{-R} = \text{const}. \quad (2.26.2)$$

A third identity, which relates p and v , may be derived from (2.25) with the aid of a differential form of the ideal gas law

$$d \ln p + d \ln v = d \ln T, \quad (2.27)$$

which follows from (1.1). Using (2.27) to eliminate T from (2.25.2) gives

$$c_v d \ln p + c_p d \ln v = 0. \quad (2.28)$$

Integrating then yields

$$p^{c_v} v^{c_p} = \text{const}. \quad (2.29)$$

The three identities (2.26.1), (2.26.2), and (2.29), may be expressed in terms of dimensionless constants as

$$T v^{\gamma-1} = \text{const} \quad (2.30.1)$$

$$T p^{-\kappa} = \text{const} \quad (2.30.2)$$

$$p v^{\gamma} = \text{const}. \quad (2.30.3)$$

Known as *Poisson's equations*, (2.30) define adiabatic paths in the state space of an ideal gas. Each describes the evolution of a state variable during an adiabatic process in terms of only one other state variable. Thus, the change of a single state variable,

together with the condition that the process be adiabatic, is sufficient to determine the change of a second state variable. That, in turn, determines the change of thermodynamic state. For this reason, an adiabatic system possesses only one independent state variable and thus only one thermodynamic degree of freedom.

Because the state space of a pure substance is represented by the plane of any two intensive properties z_1 and z_2 , a thermodynamic process describes a contour

$$g(z_1, z_2) = \text{const.}$$

Poisson's equations (2.30) are of this form. They describe a family of contours, known as *adiabats*, in the plane of any two of the state variables p , T , and v (Fig. 2.8). In similar fashion, isobaric processes describe a family of *isobars* ($p = \text{const}$), isothermal processes describe a family of *isotherms* ($T = \text{const}$), and isochoric processes describe a family of *isochores* ($v = \text{const}$), which are superposed in Fig. 2.8. Each of the latter paths is characterized by invariance of a certain state variable. The same applies to adiabats.

2.4.1 Potential temperature

Poisson's relation between pressure and temperature motivates the introduction of a new state variable, one that is preserved during an adiabatic process. The *potential temperature* θ is defined as that temperature the system would assume were it compressed or expanded adiabatically to a reference pressure of $p_0 = 1000$ hPa. According to (2.30.2), an adiabatic process from the state (p, T) to the reference state (p_0, θ) satisfies

$$\theta p_0^{-\kappa} = T p^{-\kappa}.$$

Hence, the potential temperature is described by

$$\frac{\theta}{T} = \left(\frac{p_0}{p} \right)^\kappa. \quad (2.31)$$

A function of pressure and temperature, θ is a state variable. According to (2.31) and Poisson's relation (2.30.2), θ is invariant along an adiabatic path in state space.

Adiabatic behavior of individual air parcels is a good approximation for many atmospheric applications. Above the boundary layer and outside of cloud, the time scale of heat transfer is of order 2 weeks. It is thus long compared with the characteristic time scale of motion, which influences an air parcel through changes of pressure and expansion work. For instance, vertical motion of air and accompanying changes of pressure and volume occur inside cumulus convection on a time scale of minutes to hours. Even in motions of large horizontal dimension (e.g., in sloping convection associated with synoptic weather systems), air displacements occur on a characteristic time scale of order 1 day. Thus, for a wide range of motions, the time scale for an air parcel to adjust to changes of pressure and to perform expansion work is short compared with the characteristic time scale of heat transfer.

Under these circumstances, the potential temperature of individual air parcels is approximately conserved. An air parcel descending to greater pressure experiences an increase of temperature according to (2.30.2) due to compression work that is performed on it. However, its increase of temperature is in such proportion to its increase of pressure as to preserve the parcel's potential temperature through (2.31).

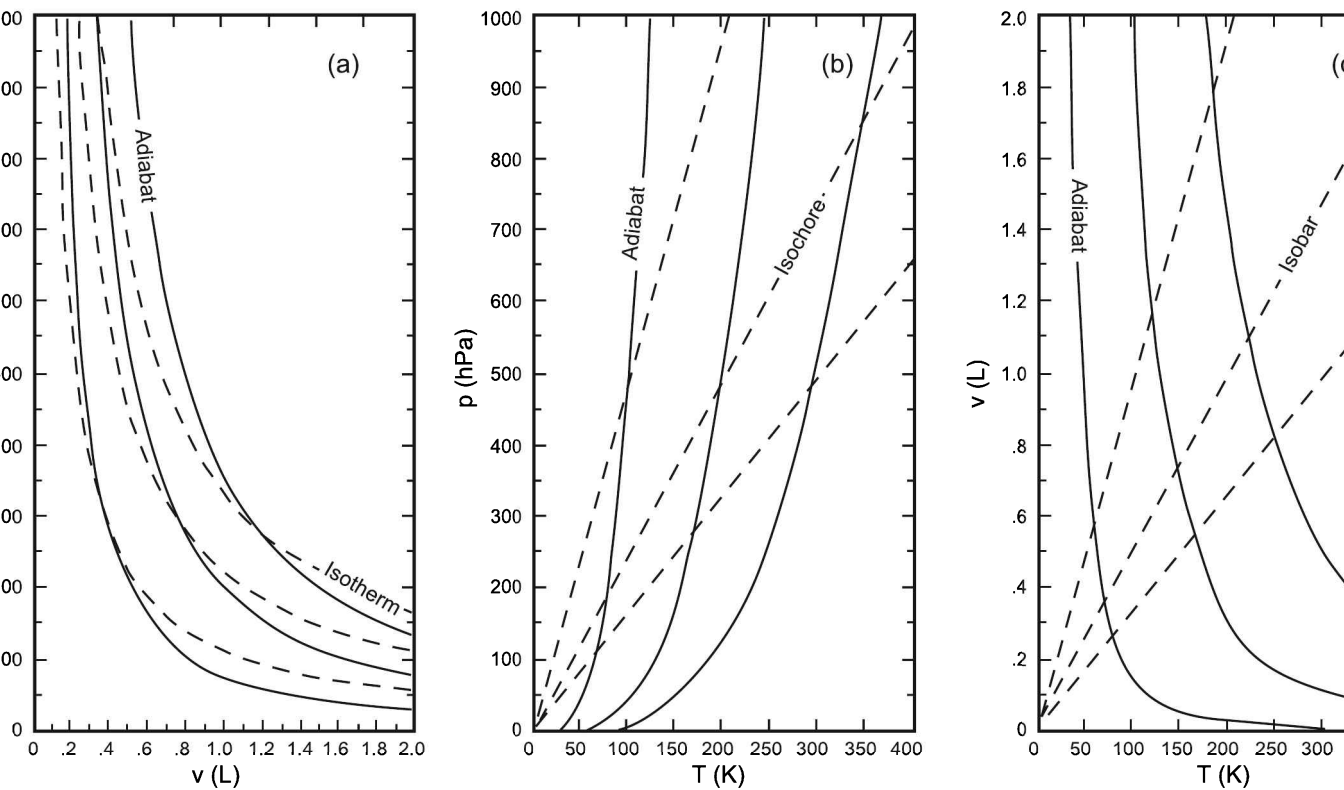


Figure 2.8 Changes of pressure, volume, or temperature in terms of one other state variable during an adiabatic process, which define *adiabats* in the state space of an ideal gas. Also shown are the corresponding changes during an isothermal process ($T = \text{const}$), an isochoric process ($v = \text{const}$), and an isobaric process ($p = \text{const}$), which define *isotherms*, *isochores*, and *isobars*, respectively.

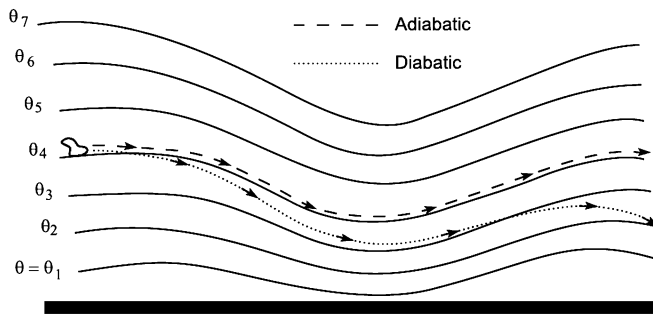


Figure 2.9 Surfaces of constant potential temperature θ . An air parcel remains coincident with a particular θ surface under adiabatic conditions (dashed), whereas it drifts across θ surfaces under diabatic conditions (dotted).

Similar considerations apply to an air parcel that is ascending. It follows that, under adiabatic conditions, θ is a conserved quantity. It therefore behaves as a tracer of air motion. On time scales for which individual parcels are adiabatic, particular values of θ track the movement of those bodies of air. Conversely, a collection of air parcels that has a particular value of θ remains coincident with the corresponding isopleth of potential temperature. The locus $\theta = \text{const}$ thus provides a constraint on how air moves.

The distribution of θ in the atmosphere is determined by the distributions of pressure and temperature. Because pressure decreases upward sharply, (2.31) implies that surfaces of constant θ tend to be quasi-horizontal like isobaric surfaces (Fig. 2.9). On time scales characteristic of air motion, a parcel initially coincident with a certain θ surface must remain coincident with that surface. Deflections of these quasi-horizontal surfaces therefore describe the vertical motion of individual bodies of air. Under steady conditions, air is constrained to move along those surfaces.

Magnified values of ozone column abundance that appear in Fig. 1.22 are attributable in part to vertical motion along θ surfaces. Contoured over Σ_{O_3} is the pressure on the 375 K potential temperature surface, which lies just above the tropopause. Above tropospheric cyclones (cf. Fig. 1.9a), that surface is deflected downward to greater pressure. Air moving along the θ surface then descends, experiencing compression. That, in turn, increases the absolute concentration ρ_{O_3} , at the expense of ozone in the surroundings, increasing Σ_{O_3} (1.25).

Another example is provided by the distribution of nitric acid on the 470 K potential temperature surface (Fig. 2.10), which also lies in the lower stratosphere. Like ozone, HNO_3 is long-lived at this level, so its mixing ratio is conserved. Therefore, r_{HNO_3} is passively rearranged by the circulation on potential temperature surfaces, along which air moves. The distribution of r_{HNO_3} on the $\theta = 470$ K surface illustrates how air is rearranged quasi-horizontally along that surface. r_{HNO_3} has been distorted into a pentagonal pattern by synoptic weather systems in the troposphere, which are distributed almost uniformly in longitude along the continuous storm track of the Southern Hemisphere (cf. Fig. 1.30a). In fact, the pattern of r_{HNO_3} suggests that air is being overturned horizontally by those synoptic disturbances (like amplifying waves in their approach to a beach).

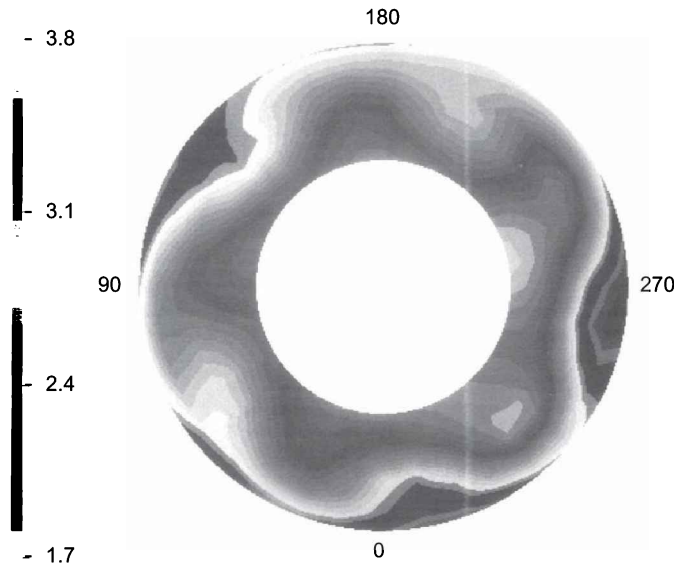


Figure 2.10 Distribution of nitric acid mixing ratio on the $\theta = 470$ K potential temperature surface over the Southern Hemisphere on January 3, 1979, as observed by Nimbus-7 LIMS. Adapted from Miles and Grose (1986).

2.4.2 Thermodynamic behavior accompanying vertical motion

According to (2.30.2), the temperature of an air parcel moving vertically changes due to expansion work. However, it does so in such proportion to its pressure as to preserve its potential temperature. An expression for the rate at which a parcel's temperature T' changes with its altitude z' under adiabatic conditions follows from (2.25.2), in combination with hydrostatic equilibrium (1.16). With the gas law, the change with altitude of environmental pressure p is described by

$$d \ln p = -\frac{dz}{H}, \quad (2.32)$$

where H is given by (1.17.2). To preserve mechanical equilibrium, the parcel's pressure p' adjusts automatically to the environmental pressure: $p' = p$. Equation (2.32) therefore also describes how the parcel's pressure changes with altitude. Then, for the parcel, (2.25.2) becomes

$$c_p dT' + g dz' = 0.$$

Thus, the temperature of a displaced air parcel evolving adiabatically decreases with its altitude at a constant rate

$$\begin{aligned} -\frac{dT'}{dz'} &= \frac{g}{c_p} \\ &= \Gamma_d, \end{aligned} \quad (2.33)$$

which defines the *dry adiabatic lapse rate*. The linear profile of temperature followed by a parcel moving vertically under adiabatic conditions (Fig. 2.11) is associated with a uniform profile of potential temperature (i.e., along which θ' for the parcel remains constant). Both describe the evolution of a displaced air parcel that does not interact

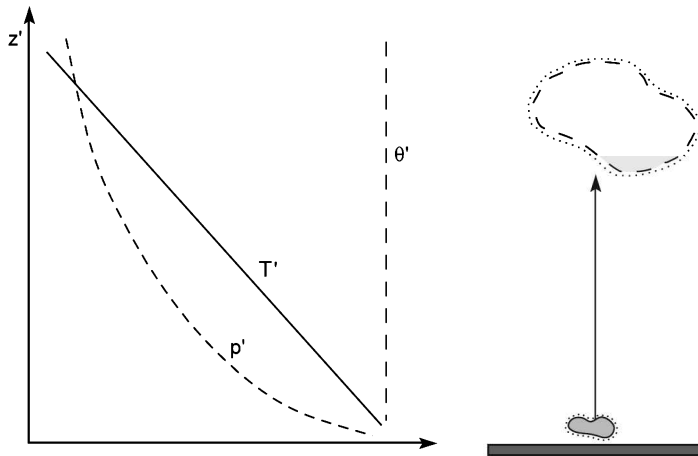


Figure 2.11 Profiles of pressure, temperature, and potential temperature for an individual air parcel ascending adiabatically.

thermally with its environment. The only interaction between the parcel and its environment is then mechanical.

Γ_d has a value of approximately 9.8 K km^{-1} . It applies to a parcel of dry air undergoing vertical motion. For reasons developed in Chap. 5, (2.33) also applies to an air parcel that is moist, but outside of cloud. Even then, heat transfer remains slow enough, compared with expansion work, to be ignored. Inside cloud, the simple behavior predicted by adiabatic considerations does not apply. It is invalidated by the release of latent heat. Accompanying the condensation of water vapor, that diabatic process occurs on the same time scale as vertical motion. Although the behavior of an air parcel is then more complex, a generalization of (2.33) applies under those circumstances (Sec. 5.4.3).

2.5 DIABATIC PROCESSES

Adiabatic conditions are violated in certain locations and over long time scales. Near the surface, thermal conduction and turbulent mixing become important because they operate on short time scales. A similar conclusion holds inside cloud, where the release of latent heat operates on the same time scale as vertical motion. For changes that occur on time scales longer than a week, radiative transfer becomes important.

Under diabatic conditions, an air parcel interacts with its environment thermally as well as mechanically. Its potential temperature is then no longer conserved. Instead, θ changes in proportion to the heat transferred into the parcel, where θ and other unprimed variables are hereafter understood to refer to an individual air parcel. Taking the logarithm of (2.31) gives a differential relation among the variables θ , p , and T

$$d \ln \theta - d \ln T = -\kappa d \ln p. \quad (2.34)$$

Likewise, dividing the First Law (2.22.2) by T and incorporating the gas law leads to a similar relation

$$d \ln T - \kappa d \ln p = \frac{\delta q}{c_p T}. \quad (2.35)$$

Combining (2.34) and (2.35) then yields

$$d \ln \theta = \frac{\delta q}{c_p T}. \quad (2.36)$$

Potential temperature increases in direct proportion to the heat transferred into the parcel (e.g., in proportion to its departure from adiabatic conditions). An air parcel will then drift across potential temperature surfaces, according to the net heat exchanged with its environment (Fig. 2.9).

Equation (2.36) may be regarded as an alternate and more compact expression of the First Law. Expansion work there has been absorbed into the state variable θ . In this light, θ for a compressible fluid, like air, is an analogue of T for an incompressible fluid, like water. Each increases in direct proportion to the heat absorbed by the system.

2.5.1 Polytropic processes

Most of the energy exchanged between the Earth's surface and the atmosphere and between one atmospheric layer and another is accomplished through radiative transfer (cf. Fig. 1.32). In fact, outside of the boundary layer and cloud, radiative transfer is the primary diabatic influence. It is sometimes convenient to model radiative transfer as a *polytropic process*

$$\delta q = c dT, \quad (2.37)$$

wherein the heat transferred into a parcel is proportional to its change of temperature. The constant of proportionality c is the *polytropic specific heat capacity*. *Newtonian cooling* damps the departure from an equilibrium temperature. It may be regarded as a polytropic process with $c < 0$. For deep temperature anomalies, as are characteristic of planetary waves, Newtonian cooling provides a convenient approximation of LW cooling to space.

For a polytropic process, the two expressions of the First Law (2.22) reduce to

$$(c_v - c)dT + p dv = 0 \quad (2.38.1)$$

$$(c_p - c)dT - v dp = 0. \quad (2.38.2)$$

Equations (2.38) resemble the First Law for an adiabatic process, but with modified specific heats. Consequently, previous formulae valid for an adiabatic process hold for a polytropic process with the transformation

$$\begin{aligned} c_p &\rightarrow (c_p - c) \\ c_v &\rightarrow (c_v - c). \end{aligned} \quad (2.39)$$

Then (2.39) may be used in (2.31) to define a *polytropic potential temperature* $\hat{\theta}$, which is conserved during a polytropic process (Prob. 2.16). Analogous to potential temperature, $\hat{\theta}$ is the temperature assumed by an air parcel if compressed polytropically to 1000 hPa. It characterizes a family of polytropes in state space.

In terms of potential temperature, the First Law becomes

$$d\ln\theta = \left(\frac{c}{c_p} \right) d\ln T. \quad (2.40)$$

If c is negative (e.g., Newtonian cooling), the effective specific heats ($c_p - c$) and ($c_v - c$) are increased from their adiabatic values. A given change of a parcel's pressure or volume in (2.38) then leads to a diminished change of its temperature. For an increase of temperature (e.g., introduced through compression), warming is diminished. The reduction from adiabatic warming reflects heat loss to the environment (e.g., through turbulent mixing or LW cooling). The heat loss is accompanied by a decrease of θ (2.40). These circumstances apply to a displaced air parcel that finds itself warmer than its environment. For a decrease of temperature (e.g., introduced through expansion), cooling is diminished. The reduction from adiabatic cooling implies heat absorption from the environment. The heat absorption is accompanied by an increase of θ . These circumstances apply to a displaced air parcel that finds itself cooler than its environment.

SUGGESTED REFERENCES

A detailed discussion of thermodynamic properties is given in *Atmospheric Thermodynamics* (1981) by Irabarne and Godson.

Statistical Thermodynamics (1973) by Lee, Sears, and Turcotte provides a clear treatment of specific heats and other gas properties from the perspective of statistical mechanics.

PROBLEMS

1. A plume of heated air leaves the cooling tower of a power plant at 1000 hPa with a temperature of 30°C. If the air may be treated as dry, to what level will the plume ascend if the ambient temperature varies with altitude as (a) $T(z) = 20 - 8z$ (°C), (b) $T(z) = 20 + z$ (°C), with z in km.
2. Suppose that the air in (1) contains moisture, but that its effect is negligible beneath a cumulus cloud, which forms overhead between 900 hPa and 700 hPa. (a) What is the potential temperature inside the plume at 950 hPa? (b) How is the potential temperature inside the plume at 800 hPa related to that at 950 hPa? (c) How is the buoyancy of air at 800 hPa related to that were the air perfectly dry?
3. One mole of water is vaporized at 100°C and 1 atm of pressure ($1.013 \cdot 10^5$ Pa), with an observed increase in volume of $3.02 \cdot 10^{-2}$ m³. (a) How much work is performed by the water? (b) How much work would be performed if the water behaved as an ideal gas and accomplished the same change of volume isobarically? (c) The enthalpy of vaporization for water is $4.06 \cdot 10^4$ J mol⁻¹. What heat input is necessary to accomplish the above change of state? Why does the heat input differ from the work performed?
4. Demonstrate (a) that $h = h(T)$, (b) relation (2.19).
5. Demonstrate that θ is conserved during an adiabatic process.
6. A large helium-filled balloon carries an instrumented payload from sea level into the stratosphere. If the balloon ascends rapidly enough for heat transfer across its surface to be negligible, what relative increase in volume must be accommodated

- for the balloon to reach 10 hPa? (Note: The balloon exerts negligible surface tension - it serves merely to isolate the helium.)
7. Through sloping convection, dry air initially at 20°C ascends from sea level to 700 hPa. Calculate (a) its initial and final specific volumes, (b) its final temperature, (c) the specific work performed, (d) changes in its specific energy and enthalpy.
 8. Commercial aircraft normally cruise near 200 hPa, where $T = -60^\circ\text{C}$. (a) Calculate the temperature air would be if compressed adiabatically to the cabin pressure of 800 hPa. (b) How much specific heat must then be added/removed to maintain a cabin temperature of 25°C?
 9. Surface wind blows down a mountain range during a *Chinook*, which is an Indian term for “snow eater.” If the temperature at 14,000' along the continental divide is -10°C and if the sky is cloud-free leeward of the divide, what is the surface temperature at Denver, which lies at approximately 5,000'?
 10. An air parcel descends from 10 hPa to 100 hPa. (a) How much specific work is performed on the parcel if its change of state is accomplished adiabatically? (b) if its change of state is accomplished isothermally at 220 K? (c) Under the conditions of (b), what heat transfer must occur to achieve the change of state? (d) What is the parcel's change of potential temperature under the same conditions?
 11. Derive (2.27) for an ideal gas.
 12. A commercial aircraft is en route to New York City from Miami, where the surface pressure was 1013 hPa and its altimeter was adjusted at takeoff to indicate 0 feet above sea level. If the surface pressure at New York is 990 hPa and the temperature is -3°C , estimate the altimeter reading when the aircraft lands if no subsequent adjustments are made.
 13. A system comprised of pure water is maintained at 0°C and 6.1 hPa, at which the specific volumes of ice, water, and vapor are: $v_i = 1.09 \cdot 10^{-3} \text{ m}^3 \text{ kg}^{-1}$, $v_w = 1.00 \cdot 10^{-3} \text{ m}^3 \text{ kg}^{-1}$, $v_v = 206 \text{ m}^3 \text{ kg}^{-1}$. (a) How much work is performed by the water substance if 1 kg of ice melts? (b) How much work is performed by the water substance if 1 kg of water evaporates? (c) Contrast the values in (a) and (b) in light of the heat transfers required to accomplish the preceding transformations of phase: $Q_{\text{fusion}} = 0.334 \cdot 10^6 \text{ J}$ and $Q_{\text{vaporization}} = 2.50 \cdot 10^6 \text{ J}$.
 14. A scuba diver releases a bubble at 30°C at a depth of 32', where the absolute pressure is 2 atm. Presume that the bubble remains intact and is large enough for heat transfer across its surface to be negligible, and that condensational heating in its interior can be ignored. Beginning from hydrostatic equilibrium and the first law, (a) derive an expression for the bubble's temperature as a function of its elevation z above its initial elevation. (b) Determine the bubble's temperature upon reaching the surface under the foregoing conditions. (c) How would these results be modified under fully realistic conditions? (Note: Normal respiration processes only about 30% of the oxygen available, so the bubble may be treated as air.)
 15. Describe physical circumstances under which polytropic heat transfer would be a useful approximation to actual heat transfer.
 16. Use the transformation (2.29) to define a *polytropic potential temperature* $\hat{\theta}$, which is conserved during a polytropic process. How is $\hat{\theta}$ related to θ ?

The Second Law and its implications

The First Law of Thermodynamics describes how the state of a system changes in response to work that it performs and heat that it absorbs. The Second Law of Thermodynamics deals with the direction of thermodynamic processes and the efficiency with which they occur. Because these characteristics control how a system evolves out of a given state, the Second Law also governs the stability of thermodynamic equilibrium.

3.1 NATURAL AND REVERSIBLE PROCESSES

A process for which the system can be restored to its initial state without leaving a net influence on the system or on its environment is said to be *reversible*. A reversible process is actually an idealization: a process that is free of friction and for which changes of state occur slowly enough for the system to remain in thermodynamic equilibrium. By contrast, a *natural process* is one that proceeds spontaneously. It can be stimulated by an infinitesimal perturbation. Because the system is then out of thermodynamic equilibrium with its surroundings, a natural process cannot be reversed entirely, namely, without leaving a net influence on either the system or its environment. A natural process is therefore inherently “irreversible.”

Irreversibility arises whenever the system is out of thermodynamic equilibrium. This occurs during rapid changes of state. Performing work across a finite pressure difference, wherein the system is out of mechanical equilibrium, is irreversible. So is transferring heat across a finite temperature difference, wherein the system is out of thermal equilibrium. A process involving friction (e.g., turbulent mixing, which damps large-scale kinetic energy) is also irreversible. So are transformations of phase when different phases of water are not in equilibrium with one another. Such phase transformations occur in states away from saturation.

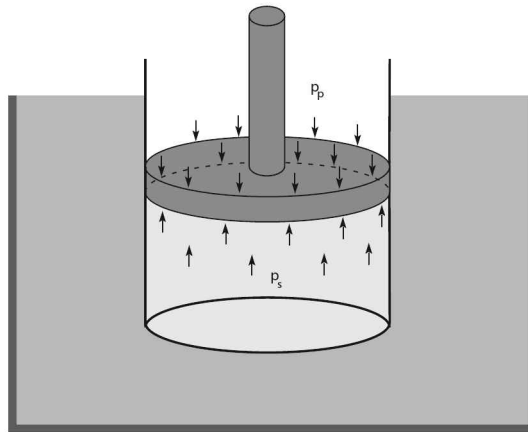


Figure 3.1 A gas at the system pressure p_s that is acted upon by a piston with pressure p_p and is maintained at constant temperature through contact with a heat reservoir.

An example of irreversible work follows from a gas that is acted upon by a piston, while being maintained at a constant temperature through contact with a heat reservoir (Fig. 3.1). Suppose the gas is first expanded isothermally at temperature T_{12} from state 1 to state 2 and subsequently restored to state 1 through isothermal compression at the same temperature (Fig. 3.2). If the cycle is executed slowly and without friction, then the system pressure p_s is uniform throughout the gas and equal to that exerted by

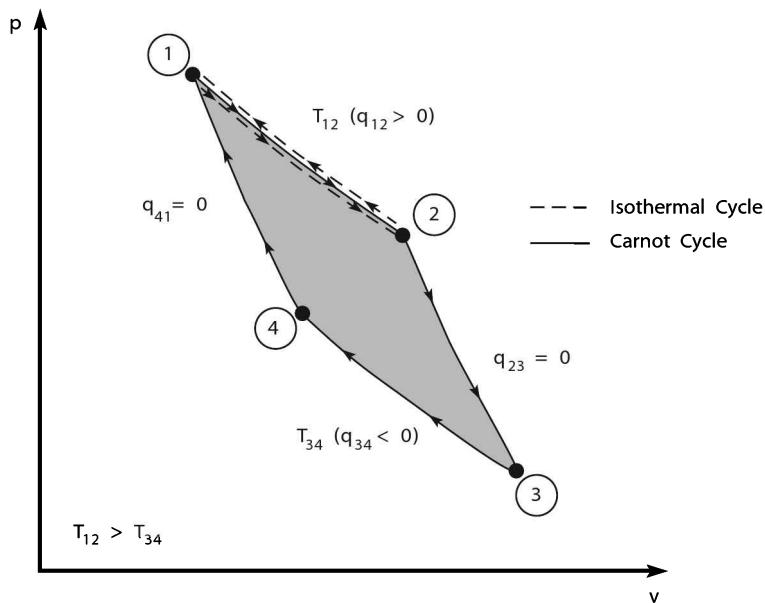


Figure 3.2 Isothermal cycle (dashed) and Carnot cycle (solid) in the $p-v$ plane.

the piston p_p at each stage of the process. The work performed by the system during isothermal expansion is thus given by

$$\begin{aligned} w_{12} &= \int_1^2 p_p dv \\ &= \int_1^2 \frac{RT_{12}}{v} dv \\ &= RT_{12} \ln \left(\frac{v_2}{v_1} \right). \end{aligned} \quad (3.1.1)$$

Likewise, the work performed on the system during isothermal compression is

$$\begin{aligned} -w_{21} &= - \int_2^1 p_p dv \\ &= RT_{12} \ln \left(\frac{v_2}{v_1} \right), \end{aligned} \quad (3.1.2)$$

equal to the work performed by the system during expansion. Hence, the net work performed during the cycle vanishes. Because $\oint du = 0$, (2.13) implies

$$\oint \delta q = \oint p dv = 0.$$

The net heat absorbed during the cycle also vanishes. Consequently, both the system and the environment are restored to their original states. The isothermal expansion from state 1 to state 2 is entirely reversible. The same holds for the isothermal compression thereafter and hence for the cycle as a whole.

If the cycle is executed rapidly, some of the gas is accelerated. Pressure is then no longer uniform across the system, so p_s does not equal p_p . During rapid compression, the piston must exert a pressure which exceeds that which occurs if the compression is executed slowly (e.g., to offset the reaction of the gas being accelerated). Therefore, the work performed on the system exceeds that performed under reversible conditions.¹ During rapid expansion, the piston must exert a pressure which is smaller than that when the expansion is executed slowly. The work performed by the system is then less than that performed under reversible conditions. Thus, executing the cycle rapidly results in

$$\oint p dv = \oint \delta q < 0.$$

Net work is performed on the system. It must be compensated by net rejection of heat to the environment.

This example illustrates that executing a transition between two states reversibly minimizes the work that must be performed on a system. Alternatively, it maximizes the work that is performed by the system. These results may be extended to a system that performs work through a conversion of heat transferred into it, viz. to a heat engine. With (2.13), it follows that the cyclic work performed by such a system is maximized when the cycle is executed reversibly. Irreversibility reduces the net work performed by the system over a cycle. By the First Law, the deficit in work must be

¹ Kinetic energy generated by the excess work is eventually dissipated and lost through heat transfer to the environment.

associated with some of the heat that is absorbed by the system being lost to the environment.

3.1.1 The Carnot cycle

Consider a cyclic process that is comprised of two isothermal legs and two adiabatic legs (Fig. 3.2). During isothermal expansion and compression, the system is maintained at constant temperature through contact with a thermal reservoir, which serves as a heat source or heat sink. During adiabatic compression and expansion, the system is thermally isolated. If executed reversibly, this process describes a *Carnot cycle*.

The leg from state 1 to state 2 is isothermal expansion at temperature T_{12} , so

$$\Delta u_{12} = 0.$$

Then the heat absorbed during that leg is given by

$$\begin{aligned} q_{12} &= w_{12} \\ &= \int_1^2 p dv \\ &= RT_{12} \ln \left(\frac{v_2}{v_1} \right). \end{aligned} \quad (3.2.1)$$

The leg from state 2 to state 3 is adiabatic, so

$$q_{23} = 0$$

and

$$\begin{aligned} -w_{23} &= \Delta u_{23} \\ &= c_v(T_{34} - T_{12}). \end{aligned} \quad (3.2.2)$$

Similarly, the leg between states 3 and 4 is isothermal compression, so the heat transfer and work are given by

$$q_{34} = w_{34} = RT_{34} \ln \left(\frac{v_4}{v_3} \right). \quad (3.2.3)$$

The closing leg between states 4 and 1 is adiabatic, so it involves work

$$w_{41} = c_v(T_{12} - T_{34}). \quad (3.2.4)$$

If the cycle is executed in reverse, the work and heat transfer during each leg are exactly opposite to those during the corresponding leg of the forward cycle. All exchanges of energy during the two cycles then cancel. The Carnot cycle is thus reversible.

During the two adiabatic legs (3.2.2) and (3.2.4), the work performed cancels. The heat transfer vanishes identically. Hence, the net work and heat transfer over the cycle follow exclusively from the isothermal legs. The volumes in (3.2.1) and (3.2.3) are related to the changes of temperature along the adiabatic legs. They follow from Poisson's identity (2.30.1)

$$\frac{T_{12}}{T_{34}} = \left(\frac{v_3}{v_2} \right)^{\gamma-1} = \left(\frac{v_4}{v_1} \right)^{\gamma-1}.$$

Thus,

$$\frac{v_2}{v_1} = \frac{v_3}{v_4}. \quad (3.3)$$

Collecting the heat transferred during the isothermal legs yields

$$\begin{aligned} \oint \delta q &= RT_{12} \ln \left(\frac{v_2}{v_1} \right) + RT_{34} \ln \left(\frac{v_4}{v_3} \right) \\ &= R(T_{12} - T_{34}) \ln \left(\frac{v_2}{v_1} \right). \end{aligned} \quad (3.4)$$

Heat transfer into the system is positive if $T_{12} > T_{34}$ and $v_2 > v_1$, namely, if isothermal expansion proceeds at high temperature and isothermal compression proceeds at low temperature. Under these conditions,

$$\oint p dv = \oint \delta q > 0. \quad (3.5)$$

The system then performs net work, which is converted from heat that it absorbs during the cycle.

The Carnot cycle is a paradigm of a heat engine. For net work to be performed by the system, heat must be absorbed at high temperature T_{12} (during isothermal expansion) and rejected at low temperature T_{34} (during isothermal compression). According to (3.2), more heat is absorbed at high temperature than is rejected at low temperature. The First Law (2.13) then implies that the net heat absorbed during the cycle is balanced by net work that is performed by the system.

If the Carnot cycle is executed in reverse, the system constitutes a refrigerator. Heat transfer into the system is negative. More heat is rejected at high temperature than is absorbed at low temperature. Then (2.13) implies that the net heat rejected during the cycle is balanced by work that must be performed on the system.

Net work and heat transfer over the Carnot cycle are proportional to the temperature difference $T_{12} - T_{34}$ between the heat source and heat sink. Not all of the heat absorbed by the system during expansion is converted into work. Some of that heat is lost during compression. The efficiency of a heat engine is therefore limited – even for a reversible cycle.

3.2 ENTROPY AND THE SECOND LAW

In the development of the First Law, we observed that, under adiabatic conditions, work is independent of path. This empirical finding permitted the introduction of a state variable, internal energy (2.11). The Second Law of Thermodynamics is inspired by the observation that, under reversible conditions, the quantity $\frac{q}{T}$ is independent of path. As for the First Law, this empirical finding permits the introduction of a state variable, *entropy*. Like internal energy, entropy is defined in terms of a quantity that, in general, is path dependent.

Consider the Carnot cycle. According to (3.4),

$$\oint \frac{\delta q}{T} = R \left[\ln \left(\frac{v_2}{v_1} \right) + \ln \left(\frac{v_4}{v_3} \right) \right] = 0.$$

This is equivalent to the identity

$$\frac{q_{12}}{T_{12}} + \frac{q_{41}}{T_{41}} = 0. \quad (3.6)$$

It turns out that this relationship holds under rather general circumstances.

Carnot's Theorem: The identity (3.6) holds for any reversible cycle between two heat reservoirs at temperatures T_{12} and T_{41} , irrespective of details of the cycle.

It can be shown that any reversible cycle can be represented as a succession of infinitesimal Carnot cycles (see, e.g., Keenan, 1970). Therefore, Carnot's theorem implies the identity

$$\oint \left(\frac{\delta q}{T} \right)_{rev} = 0, \quad (3.7)$$

irrespective of path.

By the Exact Differential Theorem (Sec. 2.1.4), it follows that the quantity $\left(\frac{\delta q}{T} \right)_{rev}$ represents a point function. That is, $\int_{(x_0, y_0)}^{(x, y)} \left(\frac{\delta q}{T} \right)_{rev}$ depends only on the thermodynamic state (x, y) , not on the path along which the system evolved to that state. We define the state variable *entropy*

$$ds = \left(\frac{\delta q}{T} \right)_{rev}, \quad (3.8)$$

which constitutes a property of the system.

Carnot's theorem and the identity (3.7) apply to a reversible process. Under more general circumstances, the following inequality holds.

Clausius Inequality: For a cyclic process,

$$\oint \frac{\delta q}{T} \leq 0, \quad (3.9)$$

where equality applies if the cycle is executed reversibly.

One of several statements of the Second Law, the Clausius inequality has the following consequences that pertain to the direction of thermodynamic processes:

- (1) Heat must be rejected to the environment somewhere during a cycle.
- (2) Under reversible conditions, more heat is absorbed at high temperature than at low temperature.
- (3) Irreversibility reduces the net heat absorbed during a cycle.

The first consequence precludes the possibility of a process that converts heat from a single source entirely into work: a *perpetual motion machine of the 2nd kind*. Some of the heat absorbed by a system that performs work must be rejected. Representing a thermal loss, that heat rejection limits the efficiency of any heat engine, even one operated reversibly. The second consequence of (3.9) implies that net work is performed by the system during a cycle (viz. it behaves as a heat engine) if heat is absorbed at high temperature and rejected at low temperature. Conversely, net work must be performed on the system during the cycle (viz. it behaves as a refrigerator) if heat is absorbed at low temperature and rejected at high temperature. The third consequence of (3.9) implies that, in the case of a heat engine, irreversibility reduces the net work

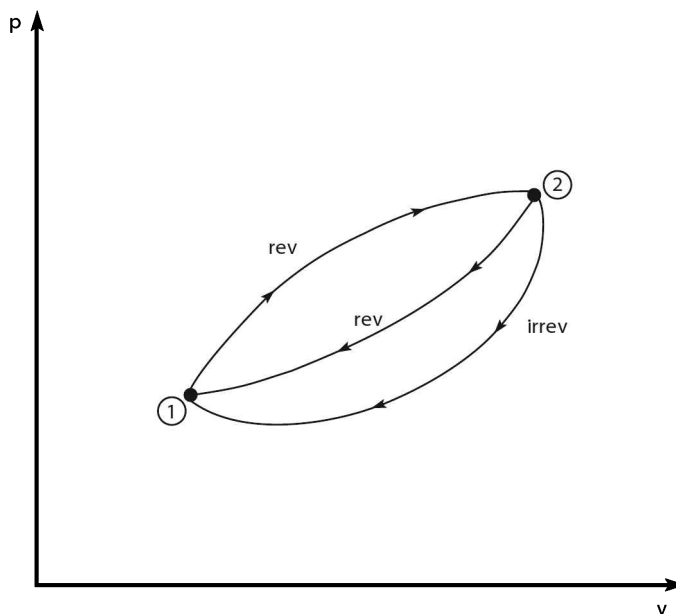


Figure 3.3 Cyclic processes between two states, consisting of (1) two reversible legs and (2) one reversible and one irreversible leg.

performed by the system. In the case of a refrigerator, irreversibility increases the net work that must be performed on the system.

A differential form of the Second Law, which applies to an incremental process, may be derived from the Clausius inequality. Consider a cycle comprised of a reversible and an irreversible leg between two states 1 and 2 (Fig. 3.3). A cycle comprised of two reversible legs may also be constructed between those states. For the first cycle, the Clausius inequality yields

$$\int_1^2 \left(\frac{\delta q}{T} \right)_{rev} + \int_2^1 \left(\frac{\delta q}{T} \right)_{irrev} \leq 0,$$

whereas for the second cycle

$$\int_1^2 \left(\frac{\delta q}{T} \right)_{rev} + \int_2^1 \left(\frac{\delta q}{T} \right)_{rev} = 0.$$

Subtracting gives

$$\int_2^1 \left(\frac{\delta q}{T} \right)_{irrev} - \int_2^1 \left(\frac{\delta q}{T} \right)_{rev} \leq 0$$

or with (3.8)

$$\int_2^1 ds \geq \int_2^1 \left(\frac{\delta q}{T} \right)_{irrev}.$$

Because it holds between two arbitrary states, the inequality must also apply to the integrands. Thus,

$$ds \geq \frac{\delta q}{T}, \quad (3.10)$$

where equality holds if the process is reversible.

Equation (3.10) is the most common form of the Second Law. In combination with (3.8), it implies an upper bound to the heat that can be absorbed by the system during a given change of state:

$$\delta q \leq T ds,$$

namely the heat absorbed when that process is executed reversibly. If the process is executed irreversibly, some heat is lost to the environment. This reduces δq and the net heat absorbed. That, in turn, reduces the net work performed by the system during a cycle (2.13).

Also represented in (3.10) is the direction of thermodynamic processes. Through the inequality, the Second Law restricts the paths along which a system can evolve. A process for which the change of entropy satisfies (3.10) is possible. If (3.10) is satisfied through equality, that process is reversible. If it is satisfied through inequality, that process is irreversible (e.g., a natural process). By contrast, a process that satisfies the reverse inequality is impossible.

3.3 RESTRICTED FORMS OF THE SECOND LAW

The entropy of a system can either increase or decrease, depending on the heat transfer needed to achieve the same change of state under reversible conditions. For certain processes, the change of entropy implied by the Second Law is simplified.

For an adiabatic process, (3.10) reduces to

$$ds_{ad} \geq 0, \quad (3.11)$$

so the entropy can then only increase. It follows that irreversible work increases a system's entropy. Letting the control surface of a hypothetical system pass to infinity eliminates heat transfer to the environment. This leads to the conclusion that the entropy of the universe can only increase. For a reversible adiabatic process, (3.11) implies that $ds = 0$. Such a process is therefore *isentropic*.

For an isochoric process, expansion work vanishes. The First Law (2.23.1) then transforms (3.8) into

$$ds = c_v \left(\frac{dT}{T} \right)_{rev}.$$

Because it involves only state variables, this expression must hold irrespective of whether the process is reversible. Hence,

$$ds_v = c_v \left(\frac{dT}{T} \right)_v. \quad (3.12)$$

In the absence of work, entropy can either increase or decrease, depending on the sign of dT . It follows that, unlike work, heat transfer can either increase or decrease a system's entropy.

According to (3.11) and (3.12), changes of entropy follow from

- (1) irreversible work
- (2) heat transfer

Irreversible work only increases s . Conversely, heat transfer (irreversible or otherwise) can either increase or decrease s . In the absence of work, the change of entropy equals $\frac{\delta q}{T}$, irrespective of path (e.g., whether or not the process is reversible).

3.4 THE FUNDAMENTAL RELATIONS

Substituting the Second Law (3.10) into the two forms of the First Law (2.22) leads to the inequalities

$$du \leq Tds - pdv \quad (3.13.1)$$

$$dh \leq Tds + vdp \quad (3.13.2).$$

For a reversible process, these reduce to

$$du = Tds - pdv \quad (3.14.1)$$

$$dh = Tds + vdp. \quad (3.14.2)$$

Because they involve only state variables, the equalities (3.14) cannot depend on path. Consequently, they must hold irrespective of whether the process is reversible. Known as the *fundamental relations*, these identities describe the change in one state variable in terms of the changes in two other state variables. Although generally valid, the fundamental relations cannot be evaluated easily under irreversible conditions. The values of p and T in (3.14) refer to the pressure and temperature “of the system.” The latter can be specified only under reversible conditions, wherein they equal “applied values.” Under irreversible conditions, they are unknown. The relationship among these variables then reverts to the inequalities (3.13), wherein p and T denote applied values, which can be specified.

It is convenient to introduce two new state variables, which are referred to as auxiliary functions. The *Helmholtz function* is defined by

$$f = u - Ts. \quad (3.15.1)$$

The *Gibbs function* is defined by

$$\begin{aligned} g &= h - Ts \\ &= u + pv - Ts. \end{aligned} \quad (3.15.2)$$

In terms of these variables, the fundamental relations become

$$df = -sdT - pdv \quad (3.16.1)$$

$$dg = -sdT + vdp. \quad (3.16.2)$$

The Helmholtz and Gibbs functions are each referred to as the *free energy* of the system: f for an isothermal process and g for an isothermal-isobaric process, because

they reflect the energy available for conversion into work under those conditions (Prob. 3.10).

3.4.1 The Maxwell Relations

Variables appearing in (3.14) and (3.16) are not entirely independent. Involving only state variables, each fundamental relation has the form of an exact differential. According to the Exact Differential Theorem (2.8), these relations can hold only if the cross derivatives of the coefficients on the right hand sides are equal. Applying that condition to (3.14) and (3.16) yields the identities

$$\begin{aligned}\left(\frac{\partial T}{\partial v}\right)_s &= -\left(\frac{\partial p}{\partial s}\right)_v \\ \left(\frac{\partial s}{\partial v}\right)_T &= \left(\frac{\partial p}{\partial T}\right)_v \\ \left(\frac{\partial T}{\partial p}\right)_s &= \left(\frac{\partial v}{\partial s}\right)_p \\ \left(\frac{\partial s}{\partial p}\right)_T &= -\left(\frac{\partial v}{\partial T}\right)_p,\end{aligned}\tag{3.17}$$

which are known as *the Maxwell relations*.

3.4.2 Noncompensated heat transfer

The inequalities in (3.13) account for additional heat rejection to the environment that occurs through irreversibility. Those inequalities may be eliminated in favor of equalities by introducing the *noncompensated heat transfer* $\delta q'$, defined by

$$\delta q = T ds - \delta q',\tag{3.18}$$

where $\delta q' > 0$. The noncompensated heat transfer represents the additional heat that is lost to the environment (and hence cannot be converted to work) as a result of irreversibility (e.g., that associated with frictional dissipation of kinetic energy).

Substituting (3.18) into the First Law (2.22.1) yields

$$du = T ds - p dv - \delta q',\tag{3.19}$$

where p and T are understood to refer to applied values. An expression for the noncompensated heat transfer can be derived from (3.14) and (3.19), which hold for applied values under reversible and irreversible conditions, respectively. Subtracting gives

$$\delta q' = (T - T_{rev}) ds - (p - p_{rev}) dv,\tag{3.20}$$

where T_{rev} and p_{rev} refer to applied values under equilibrium conditions (i.e., those assumed by the system when the process is executed reversibly). According to (3.20), noncompensated heat transfer results from (1) thermal dis-equilibrium of the system, which is represented in the difference $T - T_{rev}$, and (2) mechanical dis-equilibrium of the system, which is represented in the difference $p - p_{rev}$.

3.5 CONDITIONS FOR THERMODYNAMIC EQUILIBRIUM

By restricting the direction of thermodynamic processes, the Second Law implies whether or not a path out of a given thermodynamic state is possible. Because this determines the likelihood of the system remaining in that state, the Second Law characterizes the stability of thermodynamic equilibrium.

Consider a system in a given thermodynamic state. An arbitrary infinitesimal process out of that state is referred to as a *virtual process*. The system is said to be in *stable* or *true thermodynamic equilibrium* if no virtual process out of that state is a natural process, namely, if all virtual paths out of that state are either reversible or impossible. If all virtual paths out of the state are natural processes, the system is said to be in *unstable equilibrium*. An infinitesimal and arbitrary perturbation will then result in a finite change of state. If only some of the virtual processes out of the state are natural, the system is said to be in *metastable equilibrium*. A small perturbation then may or may not result in a finite change of state, depending on the nature of the perturbation.

These definitions may be combined with the Second Law to determine conditions that characterize thermodynamic equilibrium. The relations

$$\begin{aligned} du &\leq Tds - pdv \\ dh &\leq Tds + vdp \\ df &\leq -sdT - pdv \\ dg &\leq -sdT + vdp, \end{aligned}$$

hold for a reversible process, in the case of equality, and for an irreversible (e.g., natural) process, in the case of inequality. For a state to represent thermodynamic equilibrium, the reverse must be true, namely,

$$\begin{aligned} du &\geq Tds - pdv \\ dh &\geq Tds + vdp \\ df &\geq -sdT - pdv \\ dg &\geq -sdT + vdp, \end{aligned} \tag{3.21}$$

where inequality describes processes that are impossible. Inequalities (3.21) provide criteria for thermodynamic equilibrium. If they are satisfied by all virtual processes out of the current state, the system is stable. If only some virtual paths out of the state satisfy (3.21), the system is metastable.

Under special circumstances, simpler criteria for thermodynamic equilibrium exist. For an adiabatic enclosure, (3.10) reduces to

$$ds \geq 0,$$

where inequality corresponds to a natural process. A criterion for thermodynamic equilibrium is then

$$ds_{ad} \leq 0, \tag{3.22}$$

which describes processes that are reversible or impossible. According to (3.22), a stable state for an adiabatic system coincides with a local maximum of entropy

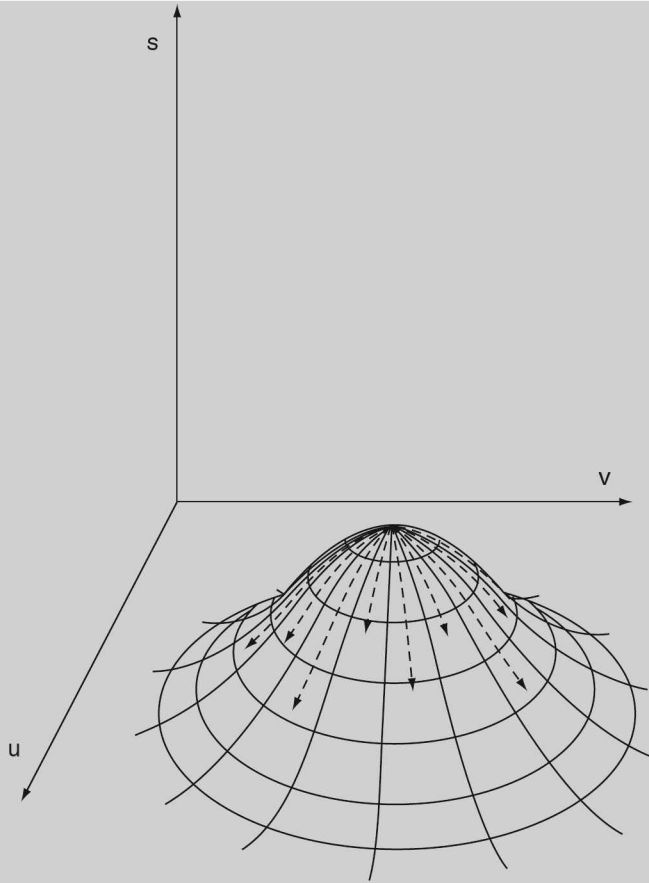


Figure 3.4 Local entropy maximum in the $u - v$ plane, symbolizing a state that corresponds to thermodynamic equilibrium.

(Fig. 3.4). Therefore, an adiabatic system's entropy must increase as it approaches thermodynamic equilibrium.

Choosing processes for which the right-hand sides of (3.21) vanish yields other criteria for thermodynamic equilibrium:

$$\begin{aligned}
 du_{s,v} &\geq 0 \\
 dh_{s,p} &\geq 0 \\
 df_{T,v} &\geq 0 \\
 dg_{T,p} &\geq 0,
 \end{aligned}
 \tag{3.23}$$

which must be satisfied by all virtual paths out of a state for it to be stable. For the processes defined, (3.23) implies that a stable state coincides with local minima in the properties u , h , f , and g , respectively. Internal energy, enthalpy, Helmholtz function, and Gibbs function must therefore decrease as a system approaches thermodynamic equilibrium.

3.6 RELATIONSHIP OF ENTROPY TO POTENTIAL TEMPERATURE

In Chap. 2, we saw that the increase of potential temperature is proportional to the heat absorbed by an air parcel. Under reversible conditions, (3.10) implies that the same is true of entropy. Substituting (2.36) into the Second Law yields

$$d \ln \theta \leq \frac{ds}{c_p}, \quad (3.24)$$

where equality holds for a reversible process. Because it involves only state variables, that equality must hold irrespective of whether the process is reversible. Hence

$$d \ln \theta = \frac{ds}{c_p}. \quad (3.25)$$

The change of potential temperature is related directly to the change of entropy. Despite its validity, (3.25) applies to properties of the system (e.g., to p and T through θ). Thus like the fundamental relations, it can be evaluated only under reversible conditions.

3.6.1 Implications for vertical motion

If a process is adiabatic, $d\theta = 0$ and $ds \geq 0$. The entropy remains constant or it can increase through irreversible work (e.g., that associated with turbulent dissipation of large-scale kinetic energy). In the case of an air parcel, the conditions for adiabatic behavior are closely related to those for reversibility. Adiabatic behavior requires not only that no heat be transferred across the control surface, but also that no heat be exchanged between one part of the system and another (see, e.g., Landau and Lifshitz 1980). The latter requirement excludes turbulent mixing, which is the principal form of mechanical irreversibility in the atmosphere. It also excludes irreversible expansion work because such work introduces internal motions that eventually result in mixing.

Because they exclude the important sources of irreversibility, the conditions for adiabatic behavior are tantamount to conditions for isentropic behavior. Thus, adiabatic conditions for the atmosphere are equivalent to requiring isentropic behavior for individual air parcels. Under these circumstances, potential temperature surfaces: $\theta = \text{const}$, coincide with isentropic surfaces: $s = \text{const}$. An air parcel coincident initially with a certain isentropic surface remains on that surface. Because those surfaces tend to be quasi-horizontal, adiabatic behavior implies no net vertical motion (cf. Fig. 2.9). Air parcels may ascend and descend along isentropic surfaces, which undulate spatially. However, they experience no systematic vertical displacement.

Under diabatic conditions, an air parcel moves across isentropic surfaces, in proportion to the heat exchanged with its environment (2.36). Consider an air parcel moving horizontally through different thermal environments, like those represented in the distribution of net radiation in Fig. 1.34. Because radiative transfer varies sharply with latitude, this occurs whenever the parcel's motion is deflected across latitude circles. Figure 3.5 shows a wavy trajectory followed by an air parcel that is initially at latitude ϕ_0 . Symbolic of the disturbed circulations in Figs. 1.9 and

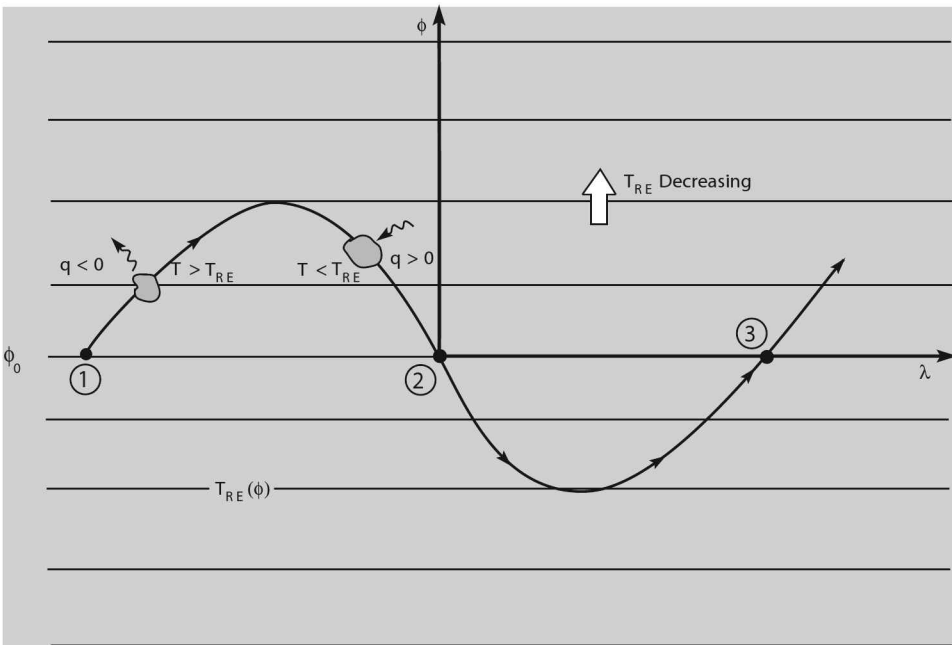


Figure 3.5 A wavy trajectory followed by an air parcel, which symbolizes the disturbed circulations in Figs. 1.9a and 1.10a. When displaced poleward, the parcel is warmer than the local radiative-equilibrium temperature, so it rejects heat. The reverse process occurs when the parcel moves equatorward.

1.10, that motion advects air through different radiative environments. Also indicated in Fig. 3.5 is the distribution of radiative-equilibrium temperature $T_{RE}(\phi)$, at which air emits radiant energy at the same rate as it absorbs it. That thermal structure is achieved if the motion is everywhere parallel to latitude circles, because air parcels then have infinite time to adjust to local thermal equilibrium.

Suppose the disturbed motion in Fig. 3.5 is sufficiently slow for the parcel to equilibrate with its surroundings at each point along the trajectory. The parcel's temperature then differs from T_{RE} only infinitesimally (Fig. 3.6), so the parcel remains in thermal equilibrium. Heat transfer along the trajectory then occurs reversibly. Between two successive crossings of the latitude ϕ_0 , the parcel absorbs heat such that

$$\int_1^2 c_p d \ln \theta = \int_1^2 \frac{\delta q}{T}. \quad (3.26)$$

If the heat exchange depends only on the parcel's temperature, e.g.,

$$\delta q = T df(T),$$

which is symbolic of radiative transfer, then (3.26) reduces to

$$c_p \ln \left(\frac{\theta_2}{\theta_1} \right) = \Delta f = 0,$$

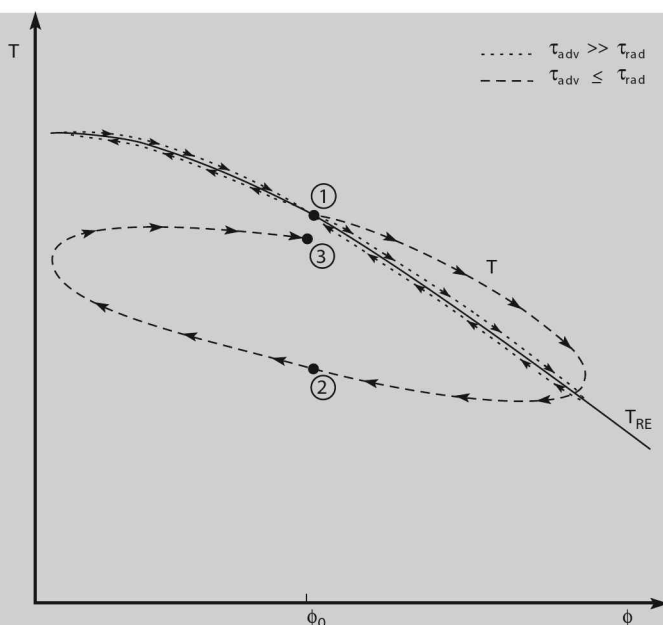


Figure 3.6 Thermal history of the idealized air parcel in Fig. 3.5. If advection occurs on a time scale much longer than radiative transfer (dotted), the parcel remains in thermal equilibrium with its surroundings (i.e., it assumes the local radiative-equilibrium temperature). The heat it absorbs and rejects is then reversible. If advection occurs on a time scale comparable to or shorter than radiative transfer (dashed), the parcel is driven out of thermal equilibrium with its surroundings. The heat it absorbs and rejects is then irreversible, which introduces a hysteresis into the parcel's thermodynamic state as it is advected through successive cycles of the disturbance (cf. Fig. 2.5).

because $T_2 = T_1 = T_{RE}(\phi_0)$. Thus $\theta_2 = \theta_1$ and the parcel is restored to its initial thermodynamic state when it returns to latitude ϕ_0 .

While moving poleward, the parcel is infinitesimally warmer than the local radiative-equilibrium temperature. It therefore emits more radiant energy than it absorbs. Rejection of heat results in the parcel drifting off its initial isentropic surface toward lower θ , which, for reasons developed in Chap. 7, corresponds to lower altitude. While moving equatorward, the parcel is infinitesimally colder than the local radiative-equilibrium temperature. It therefore absorbs more radiant energy than it emits. Absorption of heat then results in the parcel ascending to higher θ , just enough to restore the parcel to its initial isentropic surface when it returns to the latitude ϕ_0 . Successive crossings of the latitude ϕ_0 therefore result in no net vertical motion. The parcel's evolution is then perfectly cyclic.

Now suppose the motion is sufficiently fast to carry the parcel between radiative environments before it has equilibrated to the local radiative-equilibrium temperature. During the excursion poleward of ϕ_0 , the parcel is out of thermal equilibrium. Heat transfer along the trajectory therefore occurs irreversibly. Because its temperature then lags that of its surroundings, the parcel returns to the latitude ϕ_0 with

a temperature different from that initially (Fig. 3.6). By the foregoing analysis, the parcel's potential temperature θ_2 also differs from that initially

$$c_p \ln \left(\frac{\theta_2}{\theta_1} \right) = \Delta f \neq 0,$$

because $T_2 \neq T_1$. Thus the parcel is not restored to its initial isentropic surface. Rather, it remains displaced vertically after returning to the latitude ϕ_0 . Similar reasoning shows that heat transfer during the excursion equatorward of ϕ_0 (e.g., between positions 2 and 3) does not exactly cancel net heat transfer during the poleward excursion. Hence a complete cycle results in net heat transfer and therefore a net vertical displacement of the parcel from its initial isentropic surface.

Whether the parcel returns above or below that isentropic surface depends on the radiative-equilibrium temperature and on details of the motion, which control the history of heating and cooling. In either event, irreversible heat transfer introduces a hysteresis (cf. Fig. 2.5). Through it, successive cycles yield a vertical drift of air across isentropic surfaces. Because advection operates on time scales much shorter than those of radiative transfer, disturbed horizontal motion invariably drives air out of thermal equilibrium. That introduces irreversible heat transfer, which in turn drives vertical motion.

In the troposphere, irreversible heat transfer enters through quasi-horizontal mixing by synoptic weather systems. Amplifying through instability (Chap. 16), they represent exaggerated forms of the wavy pattern in Fig. 3.5. Trajectories then steepen and are eventually overturned (as in a breaking wave); cf. Figs. 2.10; 16.6. Air that is rearranged in latitude is driven out of radiative equilibrium, so it experiences irreversible heat transfer. In the stratosphere, the circulation is more often wavelike (Fig. 3.5). Through hysteresis (Fig. 3.6), it too introduces irreversible heat transfer that drives air across isentropic surfaces.² In each case, vertical motions generated in this manner contribute to mean meridional circulations, which transfer heat, moisture, and chemical constituents.

SUGGESTED REFERENCES

An illuminating treatment of the Second Law and its consequences to mechanical and chemical systems is given in *The Principles of Chemical Equilibrium* (1971) by K. Denbigh. *Statistical Physics* (1980) by Landau and Lifshitz provides a clear discussion of reversibility and its implications for fluid systems.

PROBLEMS

1. 200 g of mercury at 100°C is added to 100 g of water at 20°C. If the specific heat capacities of water and mercury are 4.18 J kg⁻¹g⁻¹ and 0.14 J kg⁻¹K⁻¹, respectively, determine (a) the limiting temperature of the mixture, (b) the change of entropy for the mercury, (c) the change of entropy for the water, (d) the change of entropy for the system as a whole.

² These sources of vertical motion are coupled to a mean poleward drift of air, which is forced by the absorption of wave activity (Sec. 18.3).

2. For reasons developed in Chap. 9, many clouds are supercooled: They contain droplets at temperatures below 0°C. Consider 1 mole of supercooled water in the metastable state: $T = -10^\circ\text{C}$ and $p = 1$ atm. Following a perturbation, the water freezes spontaneously, with the ice and its surroundings eventually returning to the original temperature. If the heat capacities of water and ice remain approximately constant with the values: $75 \text{ J K}^{-1} \text{ mol}^{-1}$ and $38 \text{ J K}^{-1} \text{ mol}^{-1}$, respectively, and if the enthalpy of fusion at 0°C is 6026 J mol^{-1} , calculate (a) the change of entropy for the water, (b) the change of entropy for its environment.
3. Consider a parcel crossing isentropic surfaces. If radiative cooling is just large enough to maintain the parcel on a fixed isobaric surface, what is the relative change of its temperature when the parcel's potential temperature has decreased to 90% of its original value?
4. Air moving inland from a cooler maritime region warms through conduction with the ground. If the temperature and potential temperature increase by 5% and 6%, respectively, determine (a) the fractional change of surface pressure between the parcel's initial and final positions, (b) the heat absorbed by the parcel.
5. During a cloud-free evening, LW heat transfer with the surface causes an air parcel to descend from 900 hPa to 910 hPa and its entropy to decrease by $15 \text{ J kg}^{-1} \text{ K}^{-1}$. If its initial temperature is 280 K, determine the parcel's (a) final temperature, (b) final potential temperature.
6. Derive expression (3.20) for the noncompensated heat transfer.
7. Air initially at 20°C and 1 atm is allowed to expand freely into an evacuated chamber to assume twice its original volume. Calculate the change of specific entropy. (Hint: How much work is performed by the air?)
8. The thermodynamic state of an air parcel is represented conveniently on the *pseudo-adiabatic chart* (Chap. 5), which displays altitude in terms of the variable p^κ . Show that the work performed during a cyclic process equals the area circumscribed by that process in the $\theta - p^\kappa$ plane.
9. One mole of water at 0°C and 1 atm is transformed into vapor at 200°C and 3 atm. If the enthalpy of vaporization for water is $4.06 \cdot 10^4 \text{ J mol}^{-1}$ and if the specific heat of vapor is approximated by

$$c_p = 8.8 - 1.9 \cdot 10^{-3} T + 2.2 \cdot 10^{-6} T^2 \quad \text{cal K}^{-1} \text{ mol}^{-1},$$

calculate the change of (a) entropy, (b) enthalpy.

10. The Helmholtz and Gibbs functions are referred to as *free energies* or *thermodynamic potentials*. Use the definitions (3.15) together with the First and Second Laws for a closed system to show that (a) under isothermal conditions, the decrease of Helmholtz function describes the maximum total work which can be performed by the system:

$$w_{\max} = -\Delta f,$$

(b) under isothermal-isobaric conditions, the decrease of Gibbs function describes the maximum work, exclusive of expansion work, which can be performed by the system:

$$w'_{\max} = -\Delta g,$$

where $\delta w' = \delta w - p dv$.

Heterogeneous systems

The thermodynamic principles developed in Chaps. 2 and 3 apply to a homogeneous system, which can involve only a single phase. For it to be in thermodynamic equilibrium, a homogeneous system must be in thermal equilibrium: At most, an infinitesimal temperature difference exists between the system and its environment. It must also be in mechanical equilibrium: At most, an infinitesimal pressure difference exists between the system and its environment.

A heterogeneous system can involve more than one phase. For such a system, thermodynamic equilibrium imposes an additional constraint. The system must also be in *chemical equilibrium*: No conversion of mass occurs from one phase to another. Analogous to thermal and mechanical equilibrium, chemical equilibrium requires a certain state variable to have, at most, an infinitesimal difference between the phases present.

4.1 DESCRIPTION OF A HETEROGENEOUS SYSTEM

For a homogeneous system, two intensive properties describe the thermodynamic state. Conversely, only two state variables may be varied independently. A homogeneous system therefore has two thermodynamic degrees of freedom. For a heterogeneous system, each phase may be regarded as a homogeneous subsystem, one that is “open” due to exchanges with the other phases present. Consequently, the number of intensive properties that describe the thermodynamic state of a heterogeneous system is proportional to the number of phases present. Were they independent, those properties would constitute additional degrees of freedom for a heterogeneous system. However, thermodynamic equilibrium between phases introduces additional constraints. They reduce the degrees of freedom of a heterogeneous system – even below those of a homogeneous system.

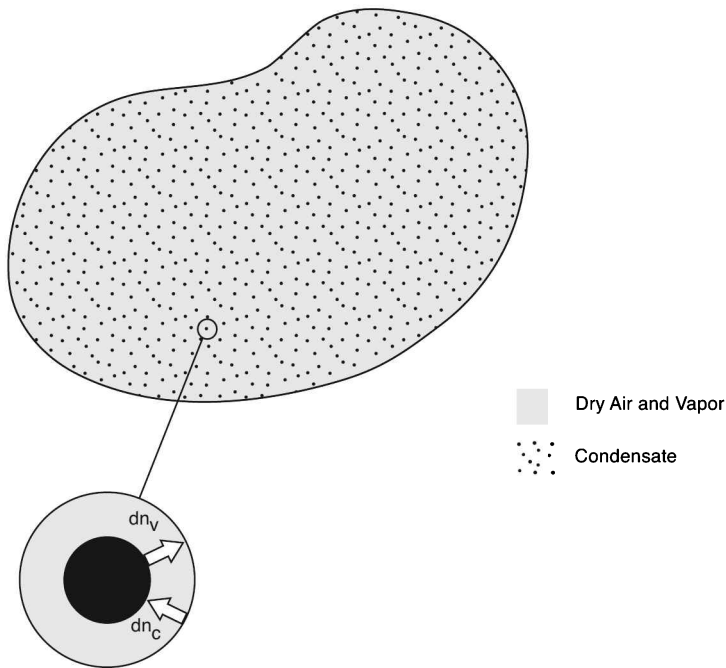


Figure 4.1 Two-component heterogeneous system of dry air and water, with the water component present in vapor and one condensed phase. The water component is transformed incrementally from condensate to vapor in amount dn_v and from vapor to condensate in amount dn_c .

The system we consider is a two-component mixture of dry air and water, with the latter existing in vapor and possibly one condensed phase (Fig. 4.1).¹ This heterogeneous system is described conveniently in terms of extensive properties, which depend on the abundance of each species present. An extensive property Z for the total system follows from the contributions from the individual phases

$$Z_{tot} = Z_g + Z_c, \quad (4.1)$$

where the subscripts g and c refer to the gas and condensate subsystems. The gas subsystem includes both dry air and vapor. The extensive property for it is specified by its pressure and temperature and the molar abundances of dry air and vapor

$$Z_g = Z_g(p, T, n_d, n_v).$$

If the condensed phase is a pure substance, the extensive property for it is determined by the pressure and temperature and the molar abundance of condensate

$$Z_c = Z_c(p, T, n_c).$$

¹ The term *vapor* refers to water in gas phase. Invisible, it should not be confused with condensate, such as aerosol or cloud.

Then changes of the extensive property Z for the individual subsystems are described by

$$dZ_g = \left(\frac{\partial Z_g}{\partial T} \right)_{pn} dT + \left(\frac{\partial Z_g}{\partial p} \right)_{Tn} dp + \left(\frac{\partial Z_g}{\partial n_d} \right)_{pTn} dn_d + \left(\frac{\partial Z_g}{\partial n_v} \right)_{pTn} dn_v \quad (4.2.1)$$

$$dZ_c = \left(\frac{\partial Z_c}{\partial T} \right)_{pn} dT + \left(\frac{\partial Z_c}{\partial p} \right)_{Tn} dp + \left(\frac{\partial Z_c}{\partial n_c} \right)_{pT} dn_c, \quad (4.2.2)$$

where the subscript n denotes holding fixed all molar values except the one being varied.

It is convenient to introduce a state variable that measures how the extensive property Z of the total system changes with an increase in one of the components (e.g., through a conversion of mass from one phase to another). It will be seen in Sec. 4.5 that a change of phase that occurs isobarically also occurs isothermally. Consequently, the foregoing state variable is expressed most conveniently for processes that occur at constant pressure and temperature. The *partial molar property* is defined as the rate an extensive property changes, under isobaric and isothermal conditions, with a change in the molar abundance of the k th species

$$\bar{Z}_k = \left(\frac{\partial Z}{\partial n_k} \right)_{pTn}, \quad (4.3.1)$$

where the subscript k refers to a particular component and phase. Similarly, the *partial specific property* is defined as

$$\bar{z}_k = \left(\frac{\partial Z}{\partial m_k} \right)_{pTm}. \quad (4.3.2)$$

In general, the partial molar and partial specific properties differ from the molar and specific properties for a pure substance:

$$\begin{aligned} \tilde{z}_k &= \frac{Z}{n_k} \\ z_k &= \frac{Z}{m_k}, \end{aligned} \quad (4.4)$$

due to interactions with the other components present.² However, for a system of dry air and water in which the latter is present only in trace abundance, such differences are small enough to be ignored. Hence, to a good approximation, the partial properties may be replaced by the molar and specific properties

$$\begin{aligned} \bar{Z}_d &\cong \tilde{z}_d & \bar{z}_d &\cong z_d \\ \bar{Z}_v &\cong \tilde{z}_v & \bar{z}_v &\cong z_v \\ \bar{Z}_c &\cong \tilde{z}_c & \bar{z}_c &\cong z_c. \end{aligned} \quad (4.5)$$

If the system is closed, the abundance of individual components is also preserved, so

$$dn_d = 0 \quad (4.6.1)$$

$$d(n_v + n_c) = 0. \quad (4.6.2)$$

² For example, adding air to a mixture of air and water at constant pressure and temperature results in a change of the extensive property Z which differs from that brought about by adding the same amount of air to a system of pure air if, in the former case, some of that air passes into solution with the water.

Adding (4.2) and incorporating (4.6) and (4.5) yields the change of the extensive property Z for the total system

$$dZ_{tot} = \left(\frac{\partial Z_{tot}}{\partial T} \right)_{pn} dT + \left(\frac{\partial Z_{tot}}{\partial p} \right)_{Tn} dp + (\bar{z}_v - \bar{z}_c) dn_v, \quad (4.7.1)$$

where

$$Z_{tot} = n_d \bar{z}_d + n_v \bar{z}_v + n_c \bar{z}_c. \quad (4.7.2)$$

In terms of specific properties, the change of Z for the total system is given by

$$dZ_{tot} = \left(\frac{\partial Z_{tot}}{\partial T} \right)_{pm} dT + \left(\frac{\partial Z_{tot}}{\partial p} \right)_{Tm} dp + (z_v - z_c) dm_v, \quad (4.8.1)$$

where

$$Z_{tot} = m_d z_d + m_v z_v + m_c z_c. \quad (4.8.2)$$

These expressions represent generalizations of the exact differential relation (2.6) for a homogeneous system to account for exchanges of mass between the phases of a heterogeneous closed system.

4.2 CHEMICAL EQUILIBRIUM

We now develop a criterion for two phases of the system to be at equilibrium with one another. In addition to thermal and mechanical equilibrium, those phases must also be in chemical equilibrium. The latter is determined by the diffusion of mass from one phase to the other, which, for reasons to become apparent, is closely related to the Gibbs function. The *chemical potential* for the k th species, μ_k , is defined as the partial molar Gibbs function:

$$\mu_k = \bar{G}_k = \left(\frac{\partial G}{\partial n_k} \right)_{pTn}, \quad (4.9)$$

which is tantamount to the molar Gibbs function \bar{g}_k . For the gas phase, (4.2.1) gives the change of Gibbs function

$$dG_g = \left(\frac{\partial G_g}{\partial T} \right)_{pn_v n_d} dT + \left(\frac{\partial G_g}{\partial p} \right)_{Tn_v n_d} dp + \mu_d dn_d + \mu_v dn_v.$$

In a constant n_v, n_d process (e.g., one not involving a phase transformation), this expression must reduce to the fundamental relation (3.16.2) for a homogeneous closed system. Accordingly, we identify

$$\begin{aligned} \left(\frac{\partial G_g}{\partial T} \right)_{pn_v n_d} &= -S_g \\ \left(\frac{\partial G_g}{\partial p} \right)_{Tn_v n_d} &= V_g. \end{aligned} \quad (4.10)$$

Because they involve only state variables, expressions (4.10) must hold irrespective of path (e.g., whether or not a phase transformation is involved). Thus,

$$dG_g = -S_g dT + V_g dp + \mu_d dn_d + \mu_v dn_v \quad (4.11)$$

describes the change of Gibbs function for the gas phase. Similar analysis leads to the relation

$$dG_c = -S_c dT + V_c dp + \mu_c dn_c \quad (4.12)$$

for the condensed phase. Adding (4.11) and (4.12) and incorporating (4.6.2) yields the change of Gibbs function for the total system

$$dG_{tot} = -S_{tot} dT + V_{tot} dp + (\mu_v - \mu_c) dn_v, \quad (4.13.1)$$

where

$$G_{tot} = n_d \bar{g}_d + n_v \bar{g}_v + n_c \bar{g}_c. \quad (4.13.2)$$

Equation (4.13.1) is a generalization to a heterogeneous system of the fundamental relation (3.16.2).

For the heterogeneous system to be in thermodynamic equilibrium, the pressures and temperatures of the different phases present must be equal. Further, there can be no conversion of mass from one phase to another. Corresponding to the fundamental relation (4.13) is the inequality

$$dG_{tot} \leq -S_{tot} dT + V_{tot} dp + (\mu_v - \mu_c) dn_v, \quad (4.14)$$

where p , T , and n_v refer to applied values and where equality and inequality correspond to reversible and irreversible processes, respectively. For the system to be in equilibrium, all virtual processes emanating from the state under consideration must be either reversible or impossible. Hence, thermodynamic equilibrium is characterized by the reverse inequality

$$dG_{tot} \geq -S_{tot} dT + V_{tot} dp + (\mu_v - \mu_c) dn_v, \quad (4.15)$$

which must hold for all virtual paths out of the state. Consider a transformation of phase that occurs at constant pressure and temperature. Then, for equilibrium,

$$dG_{tot} \geq (\mu_v - \mu_c) dn_v$$

must hold for all virtual paths. As dn_v is arbitrary, the requirement can be satisfied only if

$$\mu_v = \mu_c. \quad (4.16)$$

For chemical equilibrium, the chemical potentials of different phases of the water component must be equal. This criterion is analogous to the requirements that, under thermal and mechanical equilibrium, the temperatures and pressures of those phases must be equal.

The chemical potential determines the diffusive flux of mass from one species to another. In that respect, μ_k may be regarded as a diffusion potential from the k th species.³ Under the circumstances described by (4.16), the flux of mass from one phase of water to another is exactly balanced by a flux in the opposite sense. The net diffusion of mass therefore vanishes. If the chemical potential differs between the phases, there

³ Chemical potential is the ultimate determinant of diffusion. It supersedes concentration in certain applications; see Denbigh (1971).

will exist a net diffusion of mass from the phase of high μ to the phase of low μ . This transformation of mass will proceed until the difference of chemical potential between the phases has been eliminated, establishing chemical equilibrium.

4.3 FUNDAMENTAL RELATIONS FOR A MULTI-COMPONENT SYSTEM

In a manner similar to that used to develop the change of Gibbs function, all of the fundamental relations for a homogeneous system can be generalized to a heterogeneous system of C chemically distinct components and P phases. Expanding U , H , F , and G as above leads to the relations

$$\begin{aligned} dU &= TdS - pdV + \sum_{j=1}^P \sum_{i=1}^C \mu_{ij} dn_{ij} \\ dH &= TdS + Vdp + \sum_{j=1}^P \sum_{i=1}^C \mu_{ij} dn_{ij} \\ dU &= -SdT - pdV + \sum_{j=1}^P \sum_{i=1}^C \mu_{ij} dn_{ij} \\ dG &= -SdT + Vdp + \sum_{j=1}^P \sum_{i=1}^C \mu_{ij} dn_{ij}, \end{aligned} \quad (4.17)$$

where we have introduced alternate expressions for the chemical potential of the i th component and j th phase:

$$\mu_{ij} = \left(\frac{\partial U}{\partial n_{ij}} \right)_{SVn} = \left(\frac{\partial H}{\partial n_{ij}} \right)_{Spn} = \left(\frac{\partial F}{\partial n_{ij}} \right)_{TVn} = \left(\frac{\partial G}{\partial n_{ij}} \right)_{pTn} \quad (4.18)$$

in terms of the variables in (4.17). As will be seen shortly, the definitions in (4.18) are equivalent. However, only the last corresponds to an isobaric-isothermal process. Consequently, the first three do not represent partial molar properties according to the definition (4.3). Because transformations of phase at constant pressure also occur at constant temperature, the last expression in (4.18) is the most convenient form of chemical potential. For the same reason, the last of the fundamental relations in (4.17) is the most convenient description of such transformations.

For a closed system, the mass of each component is preserved, so

$$\sum_{j=1}^P dn_{ij} = 0, \quad i = 1, 2, 3, \dots, C. \quad (4.19)$$

For the i th component, (4.19) implies

$$\begin{aligned} \sum_{j=1}^P \mu_{ij} dn_{ij} &= \sum_{j=2}^P \mu_{ij} dn_{ij} + \mu_{i1} dn_{i1} \\ &= \sum_{j=2}^P (\mu_{ij} - \mu_{i1}) dn_{ij}, \end{aligned} \quad (4.20)$$

where $j = 1$ denotes an arbitrary reference phase for that component. Then (4.17) may be expressed

$$\begin{aligned}
 dU &= TdS - pdV + \sum_{j=2}^P \sum_{i=1}^C (\mu_{ij} - \mu_{i1}) dn_{ij} \\
 dH &= TdS + Vdp + \sum_{j=2}^P \sum_{i=1}^C (\mu_{ij} - \mu_{i1}) dn_{ij} \\
 dF &= -SdT - pdV + \sum_{j=2}^P \sum_{i=1}^C (\mu_{ij} - \mu_{i1}) dn_{ij} \\
 dG &= -SdT + Vdp + \sum_{j=2}^P \sum_{i=1}^C (\mu_{ij} - \mu_{i1}) dn_{ij},
 \end{aligned}
 \tag{4.21}$$

which represent generalizations of the fundamental relations (3.14) and (3.16) to a heterogeneous closed system. According to (4.21), each of the definitions of chemical potential in (4.18) implies the same statement of chemical equilibrium (4.16). However, only the expression involving Gibbs function applies to a process executed at constant pressure and temperature (e.g., to an isobaric transformation of phase).

4.4 THERMODYNAMIC DEGREES OF FREEDOM

Each phase of a heterogeneous system constitutes a homogeneous subsystem. Its state is therefore specified by two intensive properties. Thus, the number of intensive properties that describe the thermodynamic state of a heterogeneous system is proportional to the number of phases present.

Consider a single-component system involving two phases (e.g., pure water and vapor). The state of each homogeneous subsystem is specified by two intensive properties. Four intensive properties then specify the state of the total system. However, thermodynamic equilibrium requires each subsystem to be in thermal, mechanical, and chemical equilibrium with the other subsystem. Consequently, the heterogeneous system must also satisfy three constraints:

$$\begin{aligned}
 T_1 &= T_2 \\
 p_1 &= p_2 \\
 \mu_1 &= \mu_2.
 \end{aligned}
 \tag{4.22}$$

The heterogeneous system is left with only one independent state variable.

Thus, a one-component system involving two phases at equilibrium with one another possesses only one thermodynamic degree of freedom. Such a system must therefore possess an equation of state of the form

$$p = p(T). \tag{4.23}$$

Involving a single independent state variable, the equation of state of the heterogeneous system has the same form as that governing adiabatic changes of a homogeneous system (2.30). It describes a family of curves, along which the heterogeneous

system evolves reversibly (i.e., during which it remains in thermodynamic equilibrium). According to (4.23), fixing the temperature of a single-component mixture of two phases also fixes its pressure and vice versa. Consequently, an equilibrium transformation of phase that occurs at constant pressure also occurs at constant temperature. This feature makes the last expression in (4.18) the most useful definition of chemical potential and the change of Gibbs function the most convenient of the fundamental relations (4.21).

If all three phases are present, six intensive properties describe the state of a single-component heterogeneous system. For the system to be in thermodynamic equilibrium, those properties must also satisfy six independent constraints, like those in (4.22). Consequently, a single-component system involving all three phases possesses no thermodynamic degrees of freedom. Such a system can exist in only a single state, referred to as the *triple point*.

In general, the number of thermodynamic degrees of freedom possessed by a heterogeneous system is described by the following principle.

Gibbs' Phase Rule: The number of independent state variables for a heterogeneous system involving C chemically distinct but non-reactive components and P phases is given by

$$N = C + 2 - P. \quad (4.24)$$

According to (4.24), the degrees of freedom possessed by a heterogeneous system increases with the number of chemically distinct components but decreases with the number of phases present.⁴

4.5 THERMODYNAMIC CHARACTERISTICS OF WATER

The remainder of this chapter focuses on the thermodynamic behavior of a single-component system of pure water. This system may involve one, two, or possibly all three phases at equilibrium with one another (Fig. 4.2). Because water is a pure substance, its equation of state may, in general, be expressed

$$p = p(v, T), \quad (4.25)$$

regardless of how many phases are present (Sec 2.1.4). However, the particular form of (4.25) depends on which phases are present. If only vapor is present, (4.25) is given by the ideal gas law (1.1). If two phases are present, the equation of state must reduce to the form of (4.23).

Like any pure substance, water has an equation of state that describes a surface over the plane of two intensive properties, illustrated in Fig. 4.3 for p as a function of v and T . Portions of that surface describe states in which only a single phase is present and the system is homogeneous. In those regions of state space, the system possesses two thermodynamic degrees of freedom. Any two state variables may be varied independently. For instance, in the region of vapor, the behavior is that of an ideal gas (cf. Fig. 2.3). Both v and T are then required to specify p and the thermodynamic state.

⁴ A generalization of the phase rule to reactive components can be found in Denbigh (1971).

SINGLE-COMPONENT SYSTEM
OF
PURE WATER

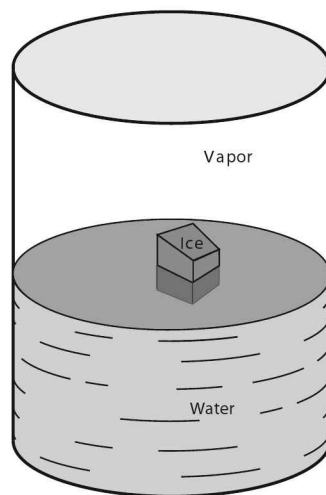


Figure 4.2 Single-component heterogeneous system of pure water.

At pressures and temperatures below the *critical point*: (p_c, T_c) , the system can assume heterogeneous states. Multiple phases then coexist at equilibrium with one another.⁵ If two phases are present, the system possesses only one thermodynamic degree of freedom (4.24). Consequently, in such regions, the surface in Fig. 4.3 assumes a simpler form. It is discontinuous in slope at the boundary with regions where the system is homogeneous. For instance, in the region of vapor and water, isotherms coincide with isobars. Specifying one of those properties determines the other and thus the thermodynamic state of the pure substance. Similar behavior is found in the regions of vapor and ice and water and ice.

At temperatures below the critical point, the system can evolve from a homogeneous state of one phase to a homogeneous state of another phase only by passing through intermediate states that are heterogeneous. Due to a change in the number of degrees of freedom, the path describing this process is discontinuous in slope at the boundary between homogeneous and heterogeneous states. Consider isobaric heat rejection (cooling), indicated in Fig. 4.3. From an initial temperature above the critical point, where the system is entirely gaseous, temperature and volume both decrease. Those changes reflect the two degrees of freedom possessed by a homogeneous system. Eventually, the process encounters the boundary separating homogeneous states, wherein only vapor is present, from heterogeneous states, wherein vapor and water coexist at equilibrium. At that state, the slope of the process changes discontinuously – because the temperature of the system can no longer decrease. Instead, isobaric cooling results in condensation of vapor and a sharp reduction of volume – all at constant temperature. This simplified behavior continues until the vapor has been converted

⁵ At temperatures above T_c , the system remains homogeneous and its properties vary smoothly across the entire range of pressure. Condensed phases are then not possible at any pressure. For those temperatures, water substance is referred to as *gas*, to distinguish it from *vapor*, which can coexist with condensed phases.

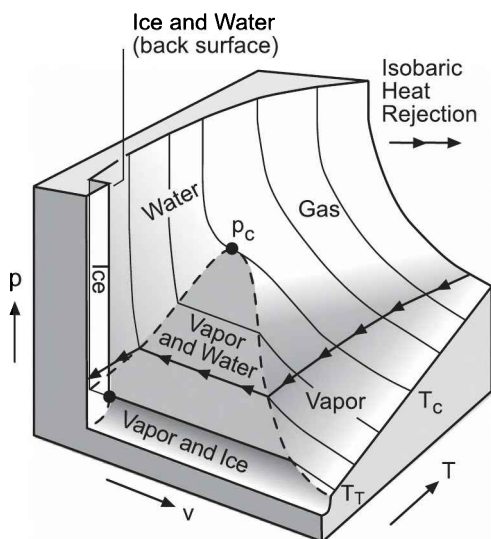


Figure 4.3 State space of a single-component system of pure water, illustrated in terms of its pressure as a function of its specific volume and temperature. Heterogeneous states occur below the *critical point* (p_c, T_c). A process corresponding to isobaric heat rejection (bold) is also indicated. In the heterogeneous region of vapor and water, isobars coincide with isotherms. Also shown is the temperature of the *triple point* (p_T, T_T, v_T), where all three phases coexist.

entirely into water. At that point, the system is again homogeneous. The slope of the process then changes discontinuously a second time. Beyond that state, heat rejection results in a decrease of both temperature and volume. Those changes reflect the two degrees of freedom that are again possessed by the system.

In a heterogeneous state, different phases coexist at equilibrium only if their temperatures, pressures, and chemical potentials are equal. The individual phases are then said to be *saturated*, because the net flux of mass from one phase to another vanishes. If one of those phases is vapor, the pressure of the heterogeneous system represents the *equilibrium vapor pressure* with respect to water or ice, denoted p_w and p_i , respectively. Should the heterogeneous system have a pressure below the equilibrium vapor pressure (e.g., below p_w), the chemical potential of the vapor will be less than that of the condensed phase. Mass will then diffuse from the condensed phase to the vapor phase, increasing the system's pressure until $p = p_w$. At that point, the net flux of mass between the two subsystems vanishes: Chemical equilibrium has been established. Conversely, a pressure above the equilibrium vapor pressure will result in a conversion of mass from vapor to condensate. Accompanied by a decrease of the system's pressure, the conversion will proceed until the difference of chemical potential between the phases has been eliminated: Chemical equilibrium is likewise established.

According to Gibbs' phase rule, there exists a single state at which all three phases coexist at equilibrium. Defined by the intersection of surfaces where water and vapor coexist, where vapor and ice coexist, and where water and ice coexist, the triple point for water is given by

$$\begin{aligned}
 p_T &= 6.1 \text{ mb} \\
 T_T &= 273 \text{ K} \\
 v_{Tv} &= 2.06 \cdot 10^5 \text{ m}^3 \text{ kg}^{-1} \\
 v_{Tw} &= 1.00 \cdot 10^{-3} \text{ m}^3 \text{ kg}^{-1} \\
 v_{Ti} &= 1.09 \cdot 10^{-3} \text{ m}^3 \text{ kg}^{-1},
 \end{aligned} \tag{4.26}$$

where the subscripts v , w , and i refer to vapor, water, and ice, respectively.

4.6 EQUILIBRIUM PHASE TRANSFORMATIONS

According to Sec. 2.3, heat transferred during an isobaric process between two homogeneous states of the same phase is proportional to the change of temperature. Under those circumstances, the constant of proportionality is the specific heat at constant pressure c_p . By contrast, heat transfer during an isobaric process between two heterogeneous states of the same two phases involves no change of temperature (Fig. 4.3). Instead, heat transfer results in a conversion of mass from one phase to the other. That transformation of phase is associated with a change of internal energy and with work being performed when the heterogeneous system's volume changes. Conversely, an isobaric transformation of phase can occur only if it is accompanied by a transfer of heat to support the change of internal energy and the work performed by the heterogeneous system during its change of volume.

4.6.1 Latent heat

The preceding effects are collected in the change of enthalpy, which equals the heat transfer into the system during an isobaric process (2.15). Analogous to the specific heat capacity at constant pressure, the specific *latent heat* of transformation is defined as the heat absorbed by the system during an isobaric transformation of phase

$$\begin{aligned} l &= \delta q_p \\ &= dh, \end{aligned} \quad (4.27.1)$$

which, by the First Law, equals the change of enthalpy during the phase transformation. In (4.27.1), the mass in l refers only to the substance undergoing the transformation of phase (e.g., to the water component in a mixture with dry air). The specific latent heats of vaporization, fusion (solid \rightarrow liquid), and sublimation (solid \rightarrow vapor) are denoted l_v , l_f , and l_s , respectively. They are related as

$$l_s = l_f + l_v. \quad (4.27.2)$$

Like specific heat capacity, latent heat is a property of the system. Thus, l depends on the thermodynamic state. For a heterogeneous system of two phases, it may be expressed $l = l(T)$. The dependence on temperature of l may be established by considering a transformation from one phase to another and how that transformation varies with temperature. Consider a homogeneous state wherein the system is entirely in phase a and another homogeneous state wherein it is entirely in phase b . By (2.23.2), an isobaric process between two homogeneous states that involve only phase a results in a change of enthalpy

$$\begin{aligned} dh_a &= \left(\frac{\partial h_a}{\partial T} \right)_p dT \\ &= c_{pa} dT. \end{aligned}$$

Likewise, an isobaric process between two homogeneous states that involve only phase b results in

$$\begin{aligned} dh_b &= \left(\frac{\partial h_b}{\partial T} \right)_p dT \\ &= c_{pb} dT. \end{aligned}$$

Subtracting gives an expression for how the difference of enthalpy between phases a and b changes with temperature

$$d(\Delta h) = (c_{pb} - c_{pa})dT.$$

Now the difference of enthalpy, at a given temperature, equals the latent heat of transformation between phase a and phase b (4.27.1). Therefore,

$$\frac{dl}{dT} = \Delta c_p, \quad (4.28)$$

where Δc_p refers to the difference of specific heat between the phases.

Known as *Kirchoff's equation*, (4.28) provides a formula for calculating the latent heat, as a function of temperature, in terms of the difference of specific heat between the two phases. Of the three latent heats, only l_f varies significantly over a range of temperature relevant to the atmosphere. At 0°C , the latent heats of water have the values

$$\begin{aligned} l_v &= 2.50 \cdot 10^6 \text{ J kg}^{-1} \\ l_f &= 3.34 \cdot 10^5 \text{ J kg}^{-1} \\ l_s &= 2.83 \cdot 10^6 \text{ J kg}^{-1}. \end{aligned} \quad (4.29)$$

Because it is defined for an isobaric process, l describes the change of enthalpy during a transformation of phase at constant pressure. The corresponding change of internal energy follows from the First Law. Consider the system in a heterogeneous state and undergoing an isobaric transformation of phase. Then (2.12.2) becomes

$$du = l - pdv. \quad (4.30)$$

For fusion, dv is negligible, so $du = l$. The change of internal energy is then just the latent heat, which also represents the change of enthalpy under those circumstances. For vaporization and sublimation,

$$dv \cong v_v, \quad (4.31)$$

where v_v refers to the volume of vapor produced. Incorporating the ideal gas law for the vapor gives the change of internal energy

$$du = l - R_v T \quad (4.32)$$

during isobaric vaporization and sublimation.

4.6.2 Clausius-Clapeyron Equation

In states involving two phases, the system of pure water possesses only one thermodynamic degree of freedom. Thus, specifying its temperature determines the system's pressure and hence its thermodynamic state. Under those circumstances, the equation of state (4.25) reduces to the form of (4.23), which describes the simplified behavior of the water surface in Fig. 4.3. The equation of state describing those heterogeneous regions may be derived from the fundamental relations, subject to conditions of chemical equilibrium.

Consider two phases a and b and a transformation between them that occurs reversibly (e.g., wherein the system remains in thermodynamic equilibrium). The heat

transfer during such a process equals the latent heat of transformation. Then, with (4.27), the Second Law becomes

$$ds = \frac{l}{T}. \quad (4.33)$$

For the system to be in chemical equilibrium, the chemical potential of phase a must equal that of phase b , so by (4.5)

$$g_a = g_b.$$

This condition must be satisfied throughout the process. Therefore, the change of Gibbs function for one phase must track that of the other

$$dg_a = dg_b.$$

Applying the fundamental relation (3.16.2) to each homogeneous subsystem then implies

$$-(s_b - s_a) dT + (v_b - v_a) dp = 0$$

or

$$\frac{dp}{dT} = \frac{\Delta s}{\Delta v}, \quad (4.34)$$

where Δ refers to the change between the phases. Incorporating (4.33) yields a relationship between the equilibrium pressure and temperature

$$\frac{dp}{dT} = \frac{l}{T\Delta v}, \quad (4.35)$$

where l is the latent heat appropriate to the phases present. Known as the *Clausius-Clapeyron equation*, (4.35) relates the equilibrium vapor pressure (e.g., $p = p_w$ or p_i) to the temperature of the heterogeneous system. It thus constitutes an equation of state for the heterogeneous system when two phases are present, describing the simplified surfaces in Fig. 4.3 that correspond to such states.

The Clausius-Clapeyron equation may be specialized to each of the heterogeneous regions in Fig. 4.3. For water and ice, l corresponds to fusion. Under those circumstances, (4.35) is expressed most conveniently in inverted form

$$\frac{dT}{dp} = \frac{T\Delta v}{l}.$$

It describes the influence on melting temperature exerted by a change of pressure. Because the change of volume during fusion is negligible, the equation of state in the region of water and ice reduces to

$$\left(\frac{dT}{dp}\right)_{fusion} \cong 0. \quad (4.36)$$

Changing the pressure has only a negligible effect on the temperature at which water is at equilibrium with ice. Consequently, the surface of water and ice in Fig. 4.3 is vertical.

For vapor and a condensed phase, l corresponds to the latent heat of vaporization or sublimation. Under those circumstances, the change of volume is approximately equal to that of the vapor produced (4.31). Thus

$$\Delta v \cong \frac{R_v T}{p},$$

which transforms (4.35) into

$$\left(\frac{d \ln p}{dT}\right)_{\substack{\text{vaporization} \\ \text{sublimation}}} = \frac{l}{R_v T^2}. \quad (4.37)$$

For $l \cong \text{const}$, (4.37) gives

$$\ln\left(\frac{p}{p_0}\right) = \frac{l}{R_v} \left(\frac{1}{T_0} - \frac{1}{T}\right). \quad (4.38)$$

Using the value of l_v in (4.29) yields the equilibrium vapor pressure with respect to water

$$\log_{10} p_w \cong 9.4041 - \frac{2.354 \cdot 10^3}{T}, \quad (4.39)$$

where p_w is in hPa and $T_0 = 0$ C is used as a reference state. Similarly, the value of l_s in (4.29) yields the equilibrium vapor pressure with respect to ice

$$\log_{10} p_i \cong 10.55 - \frac{2.667 \cdot 10^3}{T}. \quad (4.40)$$

It differs from (4.39) only modestly.

Equations (4.39) and (4.40) describe the simplified surfaces in Fig. 4.3 that correspond to vapor being in chemical equilibrium with a condensed phase (e.g., to the system pressure equaling the equilibrium vapor pressure p_w or p_i). If the system's pressure is below the equilibrium vapor pressure for the temperature of the system, water will evaporate or sublimate until the system's pressure reaches the equilibrium vapor pressure. A pressure above the equilibrium vapor pressure will result in the reverse transformation. Owing to the exponential dependence in (4.39) and (4.40), p_w and p_i vary sharply with temperature. In the presence of a condensed phase, much more water can exist in vapor phase at high temperature than at low temperature. It will be seen in Chap. 5 that the principles governing a single-component heterogeneous system carry over to a two-component system of dry air and water. The equilibrium vapor pressure then represents the maximum amount of vapor than can be supported by air at a given temperature.

The exponential dependence of p_w on T has an important implication for exchanges of water between the Earth's surface and the atmosphere. According to (4.39), warm tropical ocean with high SST can transfer much more water into the atmosphere than can colder extratropical ocean. For this reason, tropical oceans serve as the primary source of water vapor for the atmosphere. After being introduced in the tropics, water vapor is redistributed over the globe by the circulation (cf. Fig. 1.18). Much of the water vapor absorbed by the tropical atmosphere is precipitated back to the Earth's surface in organized convection inside the ITCZ (Fig. 1.30). However, latent heat that is released during condensation remains in the overlying atmosphere. Cyclic transfer of water between the ocean surface and the tropical troposphere thus results in a net transfer of heat to the atmosphere. Eventually converted into work, that heat generates kinetic energy. Along with radiative transfer from the Earth's surface (Fig. 1.32), it maintains the general circulation against frictional dissipation.

SUGGESTED REFERENCES

The Principles of Chemical Equilibrium (1971) by K. Denbigh includes a thorough treatment of phase equilibria in heterogeneous systems.

A detailed description of water substance is presented in *Atmospheric Thermodynamics* (1981) by Iribarne and Godson.

PROBLEMS

1. The Gibbs-Dalton law implies that the partial pressure of vapor at equilibrium with a condensed phase of water is the same in a mixture with dry air as it would be were the water component in isolation. Because it corresponds to the abundance of vapor at which no mass is transformed from one phase to another, this vapor pressure describes the state at which air is *saturated*. For a lapse rate of 6.5 K km^{-1} , representative of thermal structure in the troposphere (Fig. 1.2), calculate (a) the equilibrium vapor pressure as a function of altitude, and (b) the corresponding mixing ratio of water vapor as a function of altitude. (c) Apply the result to explain why water vapor is distinguished from other trace gases in Fig. 1.23.
2. A more accurate version of (4.39) is given by

$$\log_{10} p_w = -\frac{2937.4}{T} - 4.9283 \log_{10} T + 23.5471 \text{ (hPa)}.$$

- Estimate the boiling temperature for water at the altitude of (a) Denver: 5000', (b) the North America continental divide: 14,000', (c) the summit of Mt. Everest: 29,000'.
3. Present-day Venus contains little water, which is thought to have been absorbed by its atmosphere and eventually destroyed during the planet's evolution (Sec. 8.7). (a) For present conditions, wherein Venus has a surface pressure of $9 \times 10^6 \text{ Pa}$, at what surface temperature does water vapor become the atmosphere's primary constituent? Compare this with Venus' present surface temperature of 750 K. (b) Were these conditions to prevail to altitudes where energetic UV radiation is present, what would be the implication for the abundance of water on the planet? (c) Contrast these circumstances with present-day conditions on Earth.
 4. Cloud seldom forms in the stratosphere because air there is very dry. Exceptional is Polar Stratospheric Cloud (PSC), which forms over the Antarctic and, less frequently, over the Arctic. The thicker form of such cloud (Type II PSC), which is still quite tenuous compared with tropospheric cloud, is composed of ice. (a) For a mixing ratio at 80 hPa of 3 ppmv, representative of water vapor in the lower stratosphere (cf. Fig. 18.8), calculate the temperature at which ice cloud forms. (b) Referring to Fig. 1.7 and to the discussion in Sec 18.3.2, where is such cloud likely to form?
 5. The average precipitation rate inside the ITCZ is of order 10 mm day^{-1} (Fig. 9.41). (a) In W m^{-2} , calculate the average column heating rate inside the ITCZ. (b) Compare this value with the LW radiative flux emitted to the atmosphere (and largely absorbed by it) if the surface behaves as a blackbody at a temperature of 300 K (1.29).
 6. Derive the fundamental relations (4.17) for a mixture involving multiple phases.
 7. Derive alternate expressions for chemical equilibrium under (a) isothermal-isochoric conditions, (b) reversible-adiabatic and isobaric conditions.
 8. Consider a mixture of dry air and water. Describe the state space of this generally heterogeneous system and the geometry its graphical representation assumes, noting the number of thermodynamic degrees of freedom in regions where 1, 2, and 3 phases are present.
 9. Use (4.39) to estimate the latent heat of vaporization for water.

Transformations of moist air

The atmosphere is a mixture of dry air and water in varying proportion. Although its abundance varies widely, water vapor seldom represents more than a few percent of air by mass. We shall consider a two-component system comprised of these species, with water appearing in possibly one condensed phase. According to the *Gibbs-Dalton law* (which is accurate at pressures below the critical point), an individual component of a mixture of gases behaves the same as if the other components were absent. Consequently, the abundance of vapor at equilibrium with a condensed phase in a mixture of water and dry air is the same as if the water component were in isolation.¹ For this reason, concepts established in Chap. 4 for a single-component system of pure water carry over to a two-component system of dry air and water.

5.1 DESCRIPTION OF MOIST AIR

5.1.1 Properties of the gas phase

For the moment, we focus on the gas phase of this system, irrespective of whether condensate happens to be present. The vapor is in solution with dry air. It is represented by the partial pressure of vapor, e , which obeys the equation of state

$$ev_v = R_v T. \quad (5.1.1)$$

In (5.1.1),

$$v_v = \frac{V}{m_v} \quad (5.1.2)$$

¹ Implications of the Gibbs-Dalton law are developed in Keenan (1970).

is the specific volume of vapor (not to be confused with the partial volume),

$$\begin{aligned} R_v &= \frac{R^*}{M_v} \\ &= \frac{1}{\epsilon_v} R_d, \end{aligned} \quad (5.1.3)$$

and $\epsilon_v \cong 0.622$ is the ratio of molar weights defined by (1.15.2).

The *absolute humidity* $\rho_v = \frac{1}{v_v}$ measures the absolute concentration of vapor. The relative concentration of vapor is measured by the *specific humidity*

$$q = \frac{\rho_v}{\rho} = \frac{m_v}{m}, \quad (5.2)$$

which equals the ratio of the masses of vapor and mixture. Closely related is the mass *mixing ratio* $r = m_v/m_d = \rho_v/\rho_d$, which is referenced to the mass of dry air (1.14). Because

$$m = m_d + m_v, \quad (5.3)$$

the mixing ratio is approximately equal to the specific humidity

$$\begin{aligned} r &= \frac{q}{1 - q} \\ &\cong q \end{aligned} \quad (5.4)$$

because vapor is present in only trace abundance (Table 1.1). Hence, to a good approximation, q and r may be used interchangeably. Both are conserved for an individual air parcel outside regions of condensation. By contrast, measures of absolute concentration like e and ρ_v change for an individual air parcel through changes of pressure, even if the mass of vapor remains constant.

Despite the advantages of relative concentration, chemical equilibrium of the water component is controlled by the absolute concentration of vapor. For this reason, it is convenient to express the mixing ratio r in terms of the vapor pressure e . The dry air component of the mixture obeys the equation of state

$$p_d v_d = R_d T, \quad (5.5)$$

where p_d and v_d denote the partial pressure and specific volume of dry air. Dividing (5.1) by (5.5) obtains

$$\left(\frac{e}{p_d} \right) \left(\frac{v_v}{v_d} \right) = \frac{1}{\epsilon}. \quad (5.6)$$

Now

$$\begin{aligned} \left(\frac{v_v}{v_d} \right) &= \left(\frac{m_d}{m_v} \right) \\ &= \frac{1}{r}. \end{aligned}$$

Because vapor exists only in trace abundance,

$$\begin{aligned} p &= p_d + e \\ &\cong p_d. \end{aligned}$$

Then (5.6) reduces to

$$\begin{aligned}\frac{r}{\epsilon} &= \frac{e}{p} \\ &= N_v,\end{aligned}\tag{5.7}$$

where the molar fraction of vapor N_v is defined by (1.12).

Because the composition of air varies with the abundance of water vapor, so too do composition-dependent properties such as the specific gas constant R . It is convenient to absorb such variations into state variables and, instead, deal with the fixed properties of dry air. From (1.7) and (5.3), the specific gas constant of the mixture may be expressed

$$\begin{aligned}R &= (1 - q)R_d + qR_v \\ &= (1 - q)R_d + \frac{q}{\epsilon}R_d \\ &= \left[1 + \left(\frac{1}{\epsilon} - 1\right)q\right]R_d,\end{aligned}$$

or

$$R = (1 + .61q)R_d.\tag{5.8}$$

Then the equation of state for the mixture becomes

$$pv = (1 + .61q)R_d T.\tag{5.9}$$

The moisture dependence in (5.9) can be absorbed into the *virtual temperature*

$$T_v = (1 + .61q)T.\tag{5.10}$$

Then the equation of state for the gas phase of the system is simply

$$pv = R_d T_v,\tag{5.11}$$

which involves only the fixed specific gas constant of dry air. In practice, $q = O(.01)$, so, to good approximation, T may be used in place of T_v .

The specific heats of air also depend on moisture content:

$$\begin{aligned}c_v &= (1 + .97q)c_{vd} \\ c_p &= (1 + .87q)c_{pd},\end{aligned}\tag{5.12}$$

from which the dimensionless quantities γ and κ (2.21) follow directly. Like virtual temperature, c_v , c_p , γ , and κ differ only slightly from the constant values for dry air.

5.1.2 Saturation properties

Consider now the gas phase of the system in the presence of a condensed phase of water. If the vapor is in chemical equilibrium with the condensed phase, it is said to be *saturated*. Corresponding to this condition, and for a given pressure and temperature, are particular values of the foregoing moisture variables. They are referred to as *saturation values*. According to the Gibbs-Dalton law, the *saturation vapor pressure* with respect to water e_w is identical to the equilibrium vapor pressure p_w of a

single-component system of vapor and water (Sec. 4.6). Likewise, the saturation vapor pressure with respect to ice e_i is identical to p_i for a single-component system.²

The saturation vapor pressure e_c , where c denotes either of the condensed phases, is a function of temperature alone. Described by the Clausius-Clapeyron relations (4.39) and (4.40), e_c increases sharply with temperature. The *saturation specific humidity* q_c and the *saturation mixing ratio* r_c , which follows from e_c through (5.7), also describe the abundance of vapor at equilibrium with a condensed phase. Like the saturation vapor pressure, these quantities are state variables. But, because they describe the mixture and not just the vapor, q_c and r_c also depend on pressure (in accord with Gibbs' phase rule (4.24) for a two-component system involving two phases). However, the strong temperature dependence of $e_c(T)$ in the Clausius-Clapeyron equation is the dominant influence on $q_c(p, T)$ and $r_c(p, T)$. Therefore, a decrease of temperature introduced by adiabatic expansion sharply reduces the saturation values q_c and r_c . Just the reverse results from an increase of temperature introduced by adiabatic compression.

Unlike saturation values (e.g., r_c), which change with the thermodynamic state of the system, the abundance of vapor actually present (e.g., r) changes only through a transformation of phase. If no condensed phase is present (namely, under unsaturated conditions), the abundance of vapor is preserved. A decrease of temperature then results in a decrease of r_c , but no change of r . On the other hand, if the system is saturated, $r = r_c(p, T)$. A change of state in which the system remains saturated must then result in a change of both $r_c(p, T)$ and r . The vapor must adjust to maintain chemical equilibrium with the condensate present.

Two other quantities are used to describe the abundance of water vapor. The *relative humidity* is defined as

$$\begin{aligned} RH &= \frac{N_v}{N_{vc}} \\ &= \frac{e}{e_c} \approx \frac{r}{r_c}, \end{aligned} \quad (5.13)$$

where N_{vc} denotes the saturation molar abundance of vapor, and equilibrium with respect to water is implicit. The *dew point temperature* T_d is defined as that temperature to which the system must be cooled "isobarically" to achieve saturation. If saturation occurs below 0°C, that temperature is the *frost point temperature* T_f . The *dew point spread* is given by the difference $(T - T_d)$. For a given temperature T , a high dew point implies a small dew point spread. Each reflects a large abundance of vapor, which requires only a small depression of temperature to achieve saturation.

Neither relative humidity nor dew point spread are direct measures of vapor concentration. Rather, they describe how far the system is from saturation. Because saturation values increase sharply with temperature, inferring moisture content from the aforementioned quantities is misleading. For instance, r_w at 1000 hPa and 30°C is

² Strictly, the water component of the two-component system does not behave exactly as it would in isolation. Discrepancies with that idealized behavior stem from (1) near saturation, departures of the vapor from the behavior of an ideal gas; (2) the condensed phase being acted upon by the total pressure and not just that of the vapor; and (3) some of the air passing into solution with the water. However, these effects introduce discrepancies that are smaller than 1%, so they can be ignored for most applications; see Iribarne and Godson (1981) for a detailed treatment.

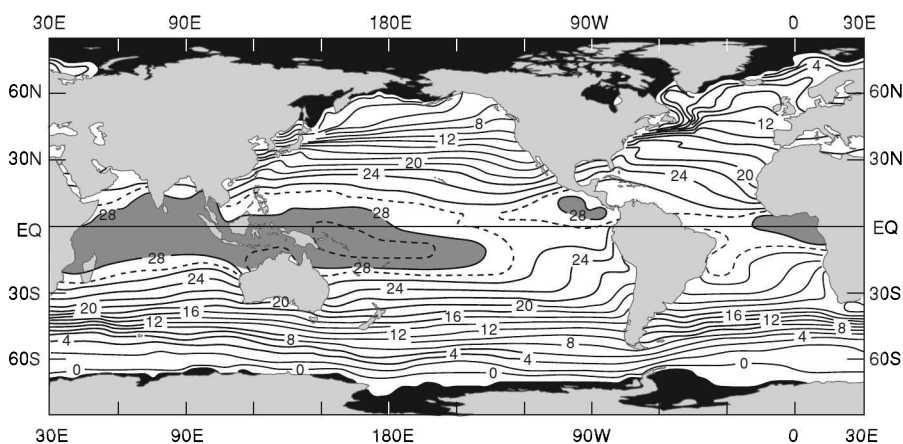


Figure 5.1 Global distribution of sea surface temperature (SST) for March. Temperatures warmer than 28°C are shaded. From Shea et al. (1990).

nearly 30 g kg^{-1} . However, at the same pressure but 0°C , it is only 4 g kg^{-1} . Thus, a relative humidity of 50% implies an abundance of vapor of 2 g kg^{-1} at 30°C but of 15 g kg^{-1} at 0°C – more than seven times greater!

5.2 IMPLICATIONS FOR THE DISTRIBUTION OF WATER VAPOR

Saturation values describe the maximum abundance of vapor that can be supported in solution with dry air at a given temperature and pressure. For that abundance, diffusion of mass from vapor to condensate is exactly balanced by diffusion of mass in the opposite sense. If the system is heterogeneous and has a vapor abundance below the saturation value (e.g., an unsaturated air parcel in contact with warm ocean surface), vapor will be absorbed until the difference of chemical potential between the phases of water has been eliminated. For this transformation to occur, the water component must absorb heat equal to the latent heat of vaporization. Conversely, if the system is heterogeneous and has a vapor abundance slightly above the saturation value (e.g., a supersaturated air parcel containing an aerosol of droplets), vapor will condense to relieve the imbalance of chemical potential. For this transformation to occur, the water component must reject heat equal to the latent heat of vaporization.

At temperatures and pressures representative of the atmosphere, the saturation vapor pressure seldom exceeds 60 hPa. The saturation mixing ratio seldom exceeds 30 g kg^{-1} or 0.030. It is for this reason that water vapor exists only in trace abundance in the atmosphere. One should note that the foregoing moisture quantities refer only to vapor – not to the total water content of the system. Condensation results in a reduction of q and r , but a commensurate increase of condensate according to (4.6.2). Unless condensate precipitates out of the system, the total water content of an air parcel is preserved.

According to the Clausius-Clapeyron equation, saturation vapor pressure increases exponentially with temperature. Thus, air can support substantially more vapor in solution at high temperature than at low temperature. For this reason, water vapor is produced efficiently in the tropics, where warm SST corresponds to a high equilibrium vapor pressure (Fig. 5.1). For the same reason, its mixing ratio decreases

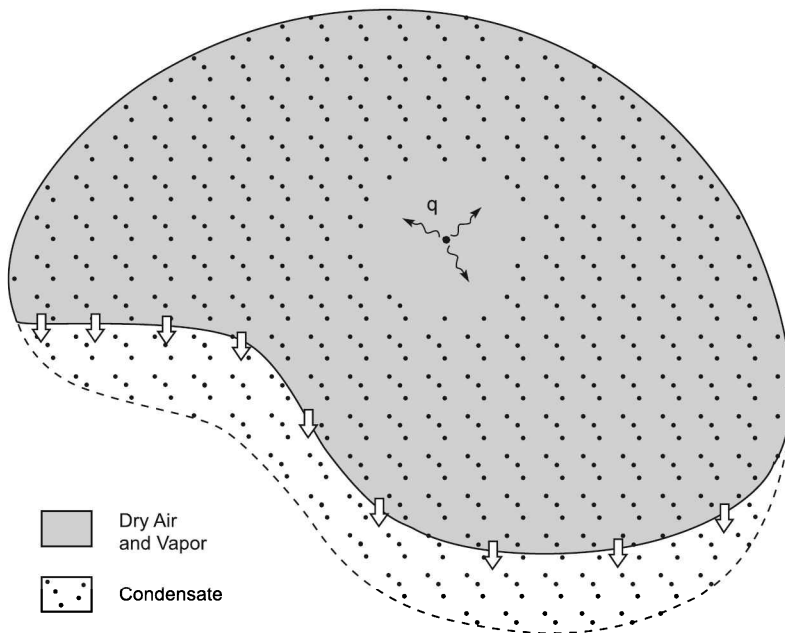


Figure 5.2 Schematic illustration of a saturated air parcel undergoing condensation, in which latent heat q is released to the gas phase and some of the condensate precipitates out of the parcel.

poleward (Fig. 1.17). Conversely, water vapor is destroyed aloft through condensation and precipitation, after displaced air parcels have cooled through adiabatic expansion work (2.33) and suffered a sharp reduction of saturation mixing ratio. Condensation also occurs at high latitude, after displaced air parcels have cooled through radiative and conductive heat transfer.

Production of vapor at an ocean surface can occur only if the water component absorbs latent heat to support the transformation of phase. Absorbed from the ocean (which then experiences evaporative cooling), that latent heat is then transferred to the air with which the vapor passes into solution. When the water re-condenses, the latent heat is released to the air surrounding the condensate (Fig. 5.2). It remains in the atmosphere after the condensate has precipitated back to the Earth's surface.³ The preceding cycle of water results in no net exchange of mass but a net transfer of heat from the ocean to the atmosphere (cf. Fig. 1.32).

5.3 STATE VARIABLES OF THE TWO-COMPONENT SYSTEM

Under unsaturated conditions, the two-component system involves only a single phase. By Gibbs' phase rule (4.24), the system then possesses three thermodynamic degrees

³ The respective heat transfer is substantial. The latent heat that is released inside a strong hurricane in one day exceeds the collective energy that is generated by the electrical grid of the United States in an entire year.

of freedom. Thus three intensive properties are required to specify the system's state if no condensate is present. Usually, the state of moist air is specified by pressure, temperature, and a humidity variable like mixing ratio that describes the abundance of vapor, e.g., $\theta = \theta(p, T, r)$. Under saturated conditions, two phases are present. One of the independent state variables is then eliminated by the constraint of chemical equilibrium. For instance, the abundance of water vapor is already determined by other state variables: $r = r_c(p, T)$. This leaves two degrees of freedom - the same number as for a homogeneous single-component system. Thus, under saturated conditions, a state variable can be specified as $\theta = \theta(p, T)$.

5.3.1 Unsaturated behavior

Under unsaturated conditions, thermodynamic processes for moist air occur much as they would for dry air because thermal properties are modified only slightly by the trace abundance of vapor. A change of temperature leads to changes of internal energy and enthalpy given by (2.18), but with slightly modified specific heats (5.12). The saturation mixing ratio r_c , which measures the capacity of air to support vapor in solution, varies with the system's pressure and temperature. By contrast, the mixing ratio r of vapor that is actually present remains constant - so long as the system is unsaturated.

The same is approximately true of the potential temperature. Under adiabatic conditions, pressure and virtual temperature change in such proportion to preserve the *virtual potential temperature*

$$\frac{\theta_v}{T_v} = \left(\frac{p_{00}}{p} \right)^{\kappa_d}, \quad (5.14)$$

where κ_d denotes the value for dry air. Like virtual temperature, θ_v is nearly identical to its counterpart for dry air, θ . Therefore, virtual properties of moist air like T_v and θ_v will hereafter be referred to by their counterparts for dry air, but the former will be understood to apply in a strict sense.

5.3.2 Saturated behavior

Under saturated conditions (e.g., in the presence of droplets), the aforementioned relationships no longer hold. Because the vapor must then be at equilibrium with a condensed phase, $e = e_c$. A change of thermodynamic state that alters e_c must then also alter e . It must therefore result in a transformation of mass from one phase of the water component to another. Accompanying that transformation of mass is an exchange of latent heat between the condensed and gas phases of the heterogeneous system (Fig. 5.2). That heat exchange alters the potential temperature of the gas phase via (2.36).

State variables describing the two-component heterogeneous system must account for these changes. For a closed system, (4.7) gives the change of total enthalpy

$$dH = \left(\frac{\partial H}{\partial T} \right)_{pm} dT + \left(\frac{\partial H}{\partial p} \right)_{Tm} dp + (h_v - h_c) dm_v. \quad (5.15)$$

Because

$$\begin{aligned}
 H &= m_d h_d + m_v h_v + m_c h_c, \\
 \left(\frac{\partial H}{\partial T}\right)_{pm} &= m_d \left(\frac{\partial h_d}{\partial T}\right)_{pm} + m_v \left(\frac{\partial h_v}{\partial T}\right)_{pm} + m_c \left(\frac{\partial h_c}{\partial T}\right)_{pm} \\
 &= m_d c_{pd} + m_v c_{pv} + m_c c_{pc}.
 \end{aligned} \tag{5.16}$$

Similarly,

$$\left(\frac{\partial H}{\partial p}\right)_{Tm} = m_d \left(\frac{\partial h_d}{\partial p}\right)_{Tm} + m_v \left(\frac{\partial h_v}{\partial p}\right)_{Tm} + m_c \left(\frac{\partial h_c}{\partial p}\right)_{Tm}.$$

For components of the gas phase, (2.18) gives

$$\left(\frac{\partial h_d}{\partial p}\right)_{Tm} = \left(\frac{\partial h_v}{\partial p}\right)_{Tm} = 0.$$

For the condensed phase, the corresponding change of enthalpy can be expressed

$$\left(\frac{\partial h_c}{\partial p}\right)_{Tm} = v_c (1 - T\alpha_p),$$

where

$$\alpha_p = \frac{1}{v} \left(\frac{\partial v}{\partial T}\right)_p$$

is the *isobaric expansion coefficient* (e.g., Denbigh, 1971). Because α_p is small for condensed phases,

$$\left(\frac{\partial h_c}{\partial p}\right)_{Tm} \cong v_c.$$

The contribution to (5.15) from this pressure term can be shown to be negligible compared with the corresponding contribution from temperature (5.16). The specific heat of the heterogeneous system is

$$c_p = \frac{m_d c_{pd} + m_v c_{pv} + m_c c_{pc}}{m}. \tag{5.17}$$

The latent heat is just the difference of enthalpy between the phases of water

$$l = h_v - h_c. \tag{5.18}$$

Incorporating the above into (5.15) then obtains the change of specific enthalpy for the system

$$\begin{aligned}
 dh &= c_p dT + l \frac{dm_v}{m} \\
 &\cong c_p dT + l dr.
 \end{aligned} \tag{5.19}$$

The last expression holds exactly if the specific enthalpy refers to a unit mass of dry air, derivation of which is left as an exercise. In (5.19), the term

$$l dr = -\delta q$$

represents the heat transferred to the gas phase (mostly dry air) from the water component when the latter undergoes condensation. It reflects an internal heat source for a saturated parcel undergoing condensation. Under adiabatic conditions, wherein heat transfer between the parcel and its environment vanishes, the First Law (2.15) reduces to

$$dh = 0.$$

Equation (5.19) then gives

$$c_p dT = -ldr.$$

A decrease in vapor leads to an increase in temperature of the parcel, which is composed chiefly of dry air.⁴

If l is treated as constant (Sec. 4.6.1), (5.19) may be integrated to yield an expression for the absolute enthalpy of the two-component heterogeneous system:

$$h = c_p T + lr + h_0, \quad (5.20)$$

where h_0 denotes the enthalpy at a suitably defined reference state.

Expressions for the internal energy and entropy of the system follow in similar fashion:

$$u = c_v T + lr + u_0 \quad (5.21)$$

$$\begin{aligned} \frac{s}{c_p} &= \ln T - \kappa_d \ln p + \frac{lr}{c_p T} + \frac{s_0}{c_p} \\ &= \ln \theta + \frac{lr}{c_p T} + \frac{s_0}{c_p}. \end{aligned} \quad (5.22)$$

Like (5.19), the expressions for absolute enthalpy, internal energy, and entropy of the two-component system are exact if referenced to a unit mass of dry air.

5.4 THERMODYNAMIC BEHAVIOR ACCOMPANYING VERTICAL MOTION

The thermodynamic state of a moist air parcel changes through vertical motion. Vertical displacement alters the parcel's environmental pressure, which varies hydrostatically according to (1.17). To preserve mechanical equilibrium, the parcel expands or contracts, which results in work being performed. Compensating that work is a change of internal energy. It alters the parcel's temperature and, hence, the saturation vapor pressure of the two-component system.

5.4.1 Condensation and the release of latent heat

From (4.38), the change of saturation vapor pressure between a reference temperature T_0 and a temperature T may be expressed

$$\ln \left(\frac{e_c}{e_{c0}} \right) = -\frac{l}{R_v} \left(\frac{1}{T} - \frac{1}{T_0} \right). \quad (5.23)$$

⁴ The effect is analogous to the combustion of fuel in a combustion engine. Like water vapor, vaporized fuel has a mixing ratio of only a few percent. Its combustion heats the dry air in which it has been dissolved, which then expands and performs work.

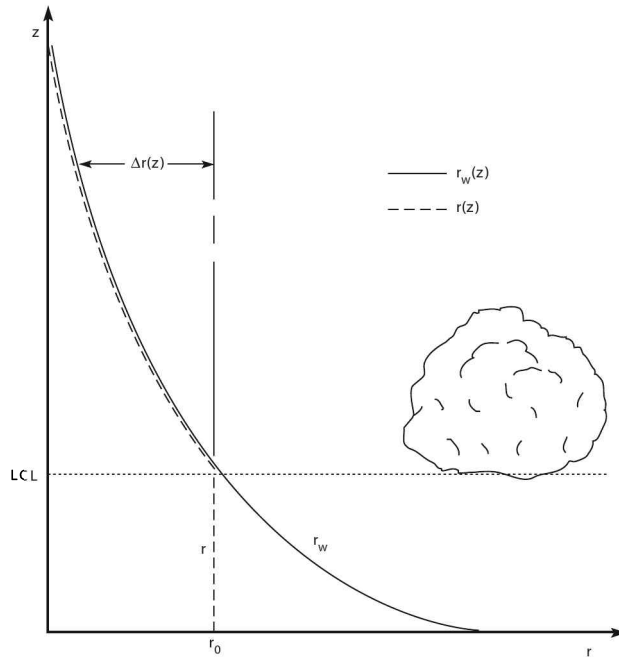


Figure 5.3 Vertical profiles of mixing ratio r and saturation mixing ratio wrt water $r_c = r_w$ for an ascending air parcel below and above the Lifting Condensation Level (LCL).

Then (5.7) implies that the saturation mixing ratio varies with pressure and temperature as

$$\frac{r_c}{r_{c0}} = \frac{e^{-\frac{L}{R_v} \left(\frac{1}{T} - \frac{1}{T_0} \right)}}{\left(\frac{p}{p_0} \right)}. \quad (5.24)$$

According to (5.24), the saturation mixing ratio increases with decreasing pressure, as accompanies upward motion. However, r_c decreases sharply with decreasing temperature, which likewise accompanies upward motion. Therefore, even though an ascending parcel's pressure decreases exponentially with its altitude, the temperature dependence in (5.24) prevails. For this reason, the parcel's saturation mixing ratio decreases monotonically with altitude.

Consider a moist air parcel ascending in thermal convection. Under unsaturated conditions, the parcel's mixing ratio and saturation mixing ratio satisfy

$$r < r_c.$$

As it rises, the parcel performs work at the expense of its internal energy. Its temperature therefore decreases at the dry adiabatic lapse rate Γ_d . From (5.24), the decrease of temperature is attended by a reduction of saturation mixing ratio r_c (Fig. 5.3). By contrast, the parcel's actual mixing ratio r remains constant, equal to its initial value $r = r_0$. The same applies to its potential temperature $\theta = \theta_0$. After sufficient upward displacement, the saturation mixing ratio will have been reduced to the actual

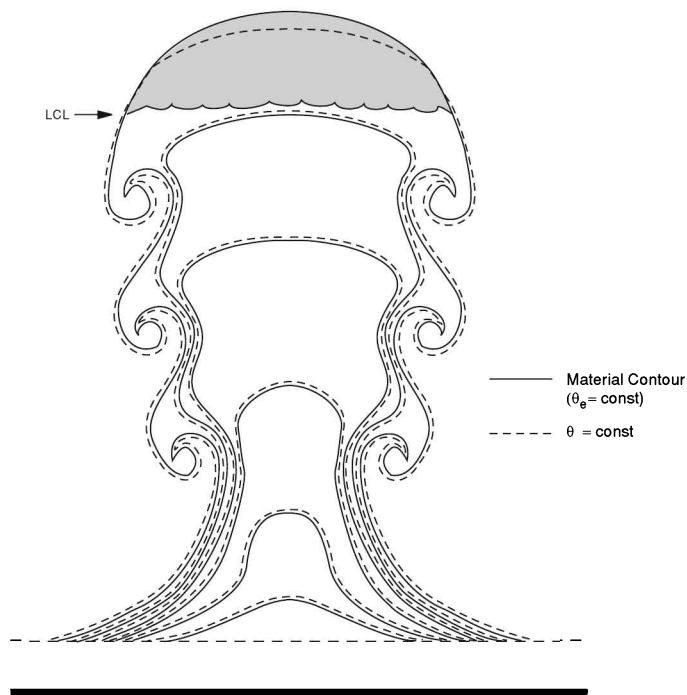


Figure 5.4 An ascending plume of moist air that develops from surface air that has become positively buoyant. Material lines (solid), which are defined by a fixed collection of air parcels, remain coincident with isentropes ($\theta = \text{const}$) below the LCL. However, they drift to higher θ and higher altitude above the LCL, where air warms through latent heat release. Although they depart from isentropes, material lines remain coincident with pseudo-isentropes ($\theta_e = \text{const}$) above the LCL.

mixing ratio:

$$r_c = r.$$

At that point, the parcel is saturated. The altitude where this first occurs is the Lifting Condensation Level (LCL). Because convective cloud forms through this process, the LCL defines the base of cumulus cloud, which is fueled by buoyant air that originates at the surface.

Below the LCL, the parcel's thermodynamic behavior may be regarded as adiabatic because the characteristic time scale for vertical displacement (e.g., from minutes inside cumulus convection to 1 day in sloping convection) is small compared with the characteristic time scale for heat transfer with the surroundings. Therefore, the parcel evolves in state space along a dry adiabat, which is described by Poisson's equations (2.30) and characterized by a constant value of θ .

In physical space, the parcel is actually part of a layer that is displaced vertically (Fig. 5.4). *Material contours*, which correspond to a fixed locus of air parcels, buckle to form a plume of moist air that ascends through positive buoyancy. Folds at the wall of the plume reflect turbulent entrainment with surrounding air. This exchange dilutes

the plume's buoyancy with that of the environment. The associated mixing introduces diabatic effects that are ignored in the present analysis. However, those effects play a key role in damping cumulus cloud (Chap. 9).⁵ Under unsaturated adiabatic conditions, θ and r are both conserved for individual parcels. Therefore, a material contour that coincides initially with a certain isentrope $\theta = \theta_0$ and mixing ratio contour $r = r_0$ remains coincident with those isopleths - so long as they reside beneath the LCL.

Above the LCL, the foregoing behavior breaks down. Continued ascent and expansional cooling reduce the saturation mixing ratio r_c below the mixing ratio r of vapor present. To restore chemical equilibrium, some of the vapor must condense - just enough to maintain chemical equilibrium:

$$r = r_c.$$

Thus, above the LCL, r and r_c both decrease with altitude (Fig. 5.3). The decrease of vapor is compensated by an increase of condensate (4.6.2).

In this fashion, ascent above the LCL wrings vapor out of solution with dry air, producing condensate (cloud droplets). The production of condensate is attended by a release of latent heat to the gas phase of the system. This heats the gas phase, which is comprised chiefly of dry air that accounts for nearly all of the parcel's mass. Latent heating thus serves as an internal heat source for the ascending parcel. Through (2.36), it increases the parcel's potential temperature. Simultaneously, it adds positive buoyancy, which reinforces ascent.

A displacement Δz above the LCL results in a decrease of water vapor mixing ratio

$$\Delta r(z) = r_c[p(z), T(z)] - r_0,$$

where z , p , and T are understood to refer to the displaced parcel with surface mixing ratio r_0 . Because r_c decreases monotonically with altitude, the greater the displacement above the LCL, the less is the vapor that remains and the greater is the condensate produced and the latent heat that has been released. Cloud droplets produced in this manner grow (through mechanisms described in Chap. 9) until they can no longer be supported by the updraft. At that point, they precipitate out of the parcel.

Because θ and r are no longer conserved, the material contour coincident initially with the isopleths $\theta = \theta_0$ and $r = r_0$ deviates from those isopleths (Fig. 5.4). The release of latent heat increases θ (2.36). Consequently, the material contour advances to isentropes of greater potential temperature. For reasons developed in Chap. 7, those isentropes lie at higher altitude than the original isentrope. Similarly, condensation reduces r . The material contour therefore moves to isopleths of smaller mixing ratio, which likewise lie at higher altitude.

Saturated air that is descending undergoes just the reverse behavior. Adiabatic compression then increases the internal energy and temperature of the gas phase. That increases the saturation mixing ratio r_c over the actual mixing ratio r . Condensate can then re-evaporate to restore chemical equilibrium: $r = r_c$. Evaporation of condensate is attended by absorption of latent heat from the gas phase. This cools the parcel, which is chiefly dry air. Evaporative cooling also introduces negative buoyancy, which reinforces

⁵ The behavior depicted in Fig. 5.4 is familiar to the sailplane pilot. The updraft fueling a growing cumulus cloud (lift) is often surrounded by a downdraft (a ring of sink), which must be penetrated to reach the core of a thermal. Above the LCL, the surrounding downdraft involves cloudy air that has been detrained from the updraft. It experiences compressional warming during its descent and evaporates.

descent. Such cooling relies on the presence of condensate that was produced during ascent. Any condensate that precipitated out of the parcel is not available to re-absorb latent heat that it released during ascent. That latent heat therefore remains in the gas phase.

The foregoing process confines water vapor to a neighborhood of the Earth's surface. Introduced over warm ocean, water vapor is extracted when moist air ascends above its LCL and experiences condensation. After cloud particles have become sufficiently large, water is removed altogether when the resulting condensate precipitates back to the Earth's surface. This process maintains upper levels of the atmosphere very dry. Even inside convective tower, mixing ratio decreases upward, because chemical equilibrium requires r to equal r_c , which decreases with altitude. By restricting vertical transport, thermodynamics prevents water vapor from reaching high altitude. There, it would be photo-dissociated by energetic radiation and eventually destroyed, when the free hydrogen produced escaped to space (Sec. 1.2.2).

Due to exchange of latent heat with the condensed phase, the gas phase of an air parcel is not adiabatic above the LCL – even though the total system may still be. Consequently, the potential temperature of the gas phase is no longer conserved. Because the parcel's mass is dominated by dry air, the change of mass during condensation exerts only a minor influence on the energetics of the gas phase. However, the accompanying release of latent heat exerts a major influence because it serves as an internal heat source for the system. Latent heat released to the gas phase during condensation offsets adiabatic cooling due to expansion work that is performed by an ascending parcel. Conversely, latent heat absorbed from the gas phase during evaporation offsets adiabatic warming due to compression work that is performed on a descending parcel. Owing to the transfer of latent heat, the parcel's temperature no longer varies with altitude at the dry adiabatic lapse rate. Instead, under saturated conditions, it varies with altitude more slowly.

5.4.2 The pseudo-adiabatic process

If expansion work occurs fast enough for heat transfer with the environment to remain negligible and if none of its moisture precipitates out, the parcel is closed. Its behavior above the LCL is then described by a reversible saturated adiabatic process. That process depends weakly on the abundance of condensate present (e.g., on how much of the system's enthalpy is represented by condensate). It therefore depends on the LCL of the parcel. However, because condensate is present only in trace abundance, its variation unnecessarily complicates the description of a saturated parcel. This complication is averted by approximating the parcel's behavior as a *pseudo-adiabatic process*: An ascending parcel is then treated as open, with condensate removed immediately after it is produced. For a descending parcel, condensate is introduced immediately before it is evaporated. Because the water component accounts for only a small fraction of the system's mass, the pseudo-adiabatic process is nearly identical to a reversible saturated adiabatic process.

A pseudo-adiabatic change of state may be constructed in two legs:

- (1) Reversible saturated adiabatic expansion (compression), which results in the production (evaporation) of condensate of mass dm_c with a commensurate release (absorption) of latent heat to (from) the gas phase.
- (2) Removal (introduction) of condensate of mass dm_c .

Because the phase transformation occurs adiabatically and reversibly, this process is isentropic

$$ds = 0.$$

As just enough vapor condenses to maintain chemical equilibrium: $r = r_c$, the change of water vapor mixing ratio is described by

$$dr = dr_c.$$

Then (5.22) implies

$$d \ln \theta = -d \left(\frac{lr_c}{c_p T} \right). \quad (5.25)$$

With the removal of condensate, (5.25) describes the change of potential temperature of the gas phase in terms of the transformation of the water component. A decrease of r_c , as accompanies ascent and expansional cooling, releases latent heat that increases θ . Conversely, an increase of r_c , as accompanies descent and compressional warming, absorbs latent heat that decreases θ .

Integrating (5.25) obtains

$$\theta e^{\left(\frac{lr_c}{c_p T} \right)} = \text{const}, \quad (5.26)$$

which describes a family of paths in the state space of saturated air. It is analogous to the family of adiabats described by Poisson's equation (2.30) for unsaturated air. The latter was used to introduce the state variable potential temperature, which is preserved during an adiabatic process. The identity (5.26) is now used to introduce another state variable, which is preserved during a pseudo-adiabatic process.

Evaluating (5.26) at a reference state of zero pressure (toward which r_c approaches zero faster than does the parcel's temperature) yields

$$\frac{\theta_e}{\theta} = e^{\left(\frac{lr_c}{c_p T} \right)}, \quad (5.27)$$

which defines the *equivalent potential temperature*. θ_e represents the maximum temperature the parcel could assume, namely, if it was raised to $z = \infty$, whereupon all of its water vapor condensed and released its latent heat, the resulting condensate was removed, and the parcel was then lowered adiabatically to the Earth's surface. According to (5.26), θ_e is constant during a pseudo-adiabatic process. Condensation is accompanied by a reduction of r_c and an increase of θ . However, the two vary in such proportion as to preserve θ_e . The same holds under evaporation.

In physical space, the material contour in Fig. 5.4 now separates from the isentrope ($\theta = \text{const}$) with which it coincided initially. Because of the release of latent heat, it advances to isentropes of higher θ . Yet, it remains coincident with the isopleth of θ_e with which it coincided initially.

Like θ , $\theta_e = \theta_e(p, T, r)$ is a state variable. θ is conserved along an adiabat in state space below the LCL. θ_e is conserved along a *pseudo* or *saturated adiabat* in state space above the LCL. An adiabatic process involving no transformation of phase is also pseudo-adiabatic. Therefore, θ_e is conserved under unsaturated conditions as well. However, the definition (5.27) can be applied only under saturated conditions because only then does $r = r_c$, as is implicit in the derivation of θ_e . Alternatively, because it is

conserved, θ_e can be calculated with r in place of r_c if T is replaced by the parcel's temperature at the LCL, where it just becomes saturated.

5.4.3 The Saturated Adiabatic Lapse Rate

The temperature of a dry air parcel decreases with its altitude at the dry adiabatic lapse rate Γ_d . To a good approximation, the same holds for a moist air parcel under unsaturated conditions - because the trace abundance of water vapor modifies thermal properties of air only slightly. Under saturated conditions, the adiabatic description of air breaks down due to the release of latent heat that accompanies the transformation of water from one phase to another. Latent heat exchanged with the gas phase then offsets adiabatic cooling and warming, which accompany ascending and descending motion.

An approximate description of how the temperature of a saturated parcel changes with altitude can be derived from the First Law with the aid of (5.19). From (2.35), the First Law for the gas phase may be expressed

$$c_p d \ln T - R d \ln p = \frac{\delta q}{T}. \quad (5.28)$$

If the parcel is unsaturated, (5.28) recovers the dry adiabatic lapse rate (2.33). If it is saturated, the heat transferred to the gas phase is given by

$$\delta q = -l dr_c. \quad (5.29)$$

Then (5.28) becomes

$$c_p d \ln T - R d \ln p = -\frac{l}{T} dr_c,$$

where l is treated as constant. With hydrostatic equilibrium (1.16), this reduces to

$$c_p dT + g dz = -l dr_c. \quad (5.30)$$

Strictly, the saturation mixing ratio depends on both pressure and temperature, so

$$dr_c = \left(\frac{\partial r_c}{\partial T} \right)_p dT + \left(\frac{\partial r_c}{\partial p} \right)_T dp.$$

However, the strong dependence on temperature conveyed from the Clausius-Clapeyron equation through (5.7) is the dominant influence on r_c . If its dependence on pressure is ignored, the change of saturation mixing ratio can be written

$$dr_c = \frac{dr_c}{dT} dT,$$

where T and z are understood to refer to the displaced parcel. Then (5.30) reduces to

$$c_p dT + g dz = -l \frac{dr_c}{dT} dT$$

or

$$\left(c_p + l \frac{dr_c}{dT} \right) dT + g dz = 0.$$

In terms of the dry adiabatic lapse rate, this may be expressed

$$\left(1 + \frac{1}{c_p} \frac{dr_c}{dT}\right) dT + \Gamma_d dz = 0. \quad (5.31)$$

Then the saturated parcel's temperature decreases with altitude according to

$$\begin{aligned} -\frac{dT}{dz} &= \frac{\Gamma_d}{1 + \frac{1}{c_p} \frac{dr_c}{dT}} \\ &= \Gamma_s, \end{aligned} \quad (5.32)$$

which defines the *saturated adiabatic lapse rate*.

Unlike Γ_d , Γ_s varies with the parcel's altitude as a result of the nonlinear dependence on T of r_c . However, because $\frac{dr_c}{dT} > 0$, (5.32) implies

$$\Gamma_s < \Gamma_d. \quad (5.33)$$

A parcel's temperature decreases with altitude slower under saturated conditions than under unsaturated conditions. This property of saturated vertical motion follows from the release of latent heat to the gas phase, which offsets cooling associated with adiabatic expansion. Although variable, the saturated adiabatic lapse rate has a value of $\Gamma_s \cong 5.5 \text{ K km}^{-1}$ for conditions representative of the troposphere. This is only slightly smaller than the global-mean lapse rate of the troposphere (Fig. 1.2). No accident, the resemblance follows from dynamical processes that are developed in Chaps. 6 and 7.

5.5 THE PSEUDO-ADIABATIC CHART

Thermodynamic processes associated with vertical motion are represented conveniently on a diagram of the state space of moist air. Shown in Fig. 5.5 is the *pseudo-adiabatic chart*, which displays, as functions of T and $-p^\kappa$,

- (1) Adiabats: $\theta = \text{const}$ (heavy solid),
- (2) Pseudo Adiabats: $\theta_e = \text{const}$ (heavy dashed), and
- (3) Isoleths of Saturation Mixing Ratio $r_w = \text{const}$ (light solid).⁶

The ordinate $-p^\kappa$ reflects altitude (1.17), which is indicated at the right for the US Standard Atmosphere. Adiabats, along which a parcel evolves under unsaturated conditions and its temperature varies with altitude at the dry adiabatic lapse rate Γ_d , appear as straight lines in this representation (2.34). Pseudo-adiabats, along which a parcel evolves under saturated conditions and its temperature varies with altitude at the saturated adiabatic lapse rate Γ_s , are curved - but only weakly. Consistent with (5.33), pseudo-adiabats ($\theta_e = \text{const}$) have a slope with respect to altitude that is everywhere smaller than that of adiabats ($\theta = \text{const}$). Both are labeled in degrees K, which correspond to the constant values of θ and θ_e that characterize those paths.

⁶ If ice is present, moisture properties actually depend on r_i . However, according to (4.39) and (4.40), isopleths of r_i differ from those of r_w only modestly.

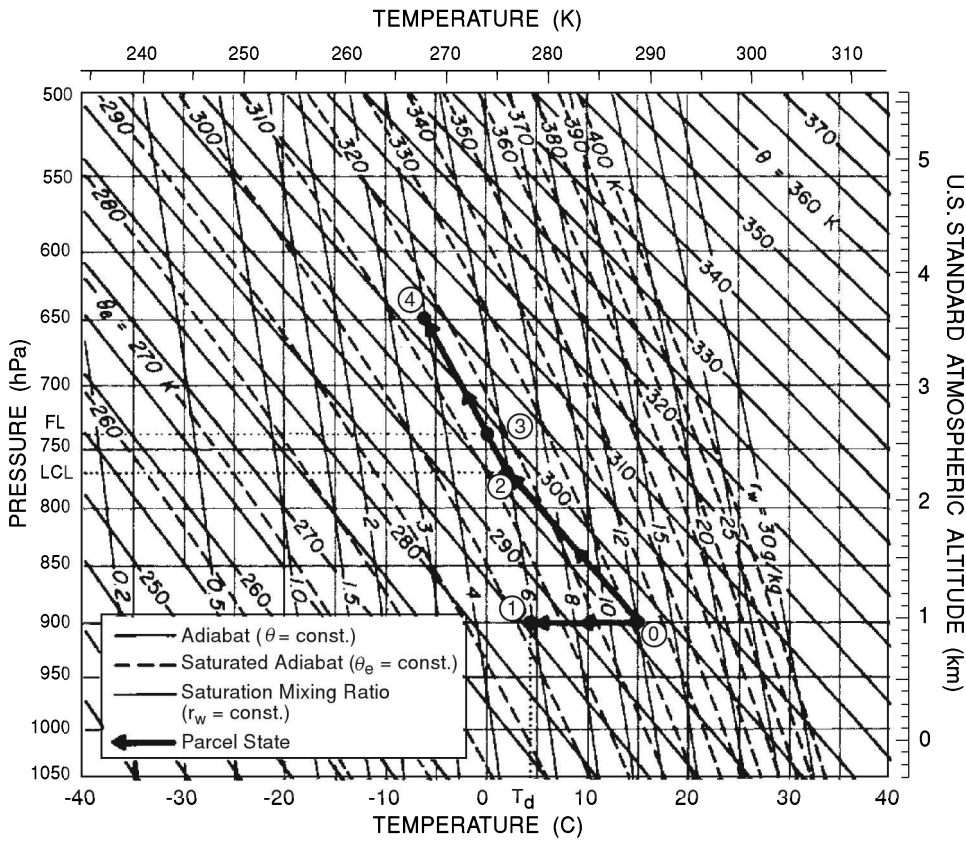


Figure 5.5 Pseudo-adiabatic chart illustrating thermodynamic processes for an ascending air parcel. Shown as functions of temperature and elevation are (1) adiabats (heavy solid), which are labeled by the constant values of θ characterizing those lines; (2) pseudo-adiabats (heavy dashed), which are labeled by the constant values of θ_e characterizing those curves; and (3) isopleths of saturation mixing ratio (light solid), which are labeled by the constant values of r_w defining those curves.

The use of the pseudo-adiabatic chart is best illustrated with an example. Consider conditions leading to the formation of a cumulus cloud. The cloud is fueled by moist air that ascends in a buoyant thermal, which is driven by surface heating (Fig. 5.4). At the surface, which is located at 900 hPa, air has a temperature of $T_0 = 15^\circ\text{C}$ and a mixing ratio $r_0 = 6.0 \text{ g kg}^{-1}$. Because air inside the thermal originates at the surface, thermodynamic properties inside the cloud, as well as those beneath the cloud, can be inferred from the evolution of an air parcel that is initially at the ground. Once the parcel's initial state (state 0) has been located on the pseudo-adiabatic chart, individual properties follow by allowing the system to evolve along certain paths in the state space of the parcel (indicated in Fig. 5.5).

Surface relative humidity

The saturation mixing ratio at the surface follows from the isopleth of r_w passing through state 0, which gives

$$r_{w0} = 12 \text{ g kg}^{-1}.$$

The initial relative humidity of surface air is thus

$$RH_0 = \frac{6}{12} = 50\%.$$

Surface potential temperature

The parcel's initial potential temperature is determined by the adiabat passing through state 0, which is characterized by the constant value

$$\theta = \theta_0 = 297 \text{ K}.$$

Surface dew point

The dew point of surface air follows from isobaric cooling out of state 0. During that process, the parcel's state evolves along the isobar $p = 900 \text{ hPa}$, which crosses isopleths of saturation mixing ratio toward lower values of r_w . Eventually, state 1 is reached, at which $r_w = r = 6 \text{ g kg}^{-1}$. The parcel is then saturated. The temperature at which this occurs defines the dew point temperature of surface air

$$T_{d0} = 4 \text{ C}.$$

The above process is responsible for the formation of ground fog. The dew point spread of surface air, 11°C in this example, provides an indirect measure of the base of convective cloud. It reflects the amount of adiabatic cooling necessary to achieve saturation (Prob. 5.26).

Cumulus cloud base

The base of convective cloud corresponds to the LCL of surface air. The latter may be determined by displacing the parcel upward adiabatically. During that process, the parcel's state evolves along the adiabat passing through state 0 ($\theta = 297 \text{ K}$). It too crosses isopleths of saturation mixing ratio toward lower values of r_w . Eventually, state 2 is reached, at which $r_w = r$ and the parcel is again saturated. The altitude where this occurs corresponds to the LCL

$$p_{LCL} = 770 \text{ hPa}.$$

The parcel's temperature at this altitude is 13 K colder than its initial temperature. However, its temperature is only 2 K colder than its initial dew point temperature. (It equals the parcel's dew point temperature at this altitude.) The small change of T_d between the surface and LCL reflects the weak pressure dependence of r_c .

Equivalent potential temperature at the surface

The equivalent potential temperature is determined once the LCL is located. Through state 2 (and along subsequent states) passes a pseudo adiabat. It defines θ_e for the parcel. Because θ_e is conserved under both saturated and unsaturated conditions, that value is also the equivalent potential temperature of the parcel below the LCL, so

$$\theta_{e0} = 315 \text{ K}$$

at 900 hPa. Note that, even though it is conserved throughout, θ_e must be inferred at and above the LCL, for reasons discussed in Sec. 5.4.2.

Freezing level of surface air

At the LCL, the parcel's temperature is greater than 0°C . Consequently, the freezing level lies higher, inside the cloud. Lower portions of the cloud contain water droplets,

whereas higher portions contain ice particles.⁷ Conditions inside the cloud may be determined by displacing the parcel above the LCL. The parcel's state then evolves along the pseudo adiabat passing through state 2. From there, the $\theta_e = 315$ K pseudo adiabat crosses adiabats toward higher θ and isopleths of saturation mixing ratio toward lower r_w . Eventually, it reaches state 3, where the temperature is 0°C . Defining the freezing level, that condition is achieved at

$$p_{FL} = 740 \text{ hPa.}$$

Liquid water content at the freezing level

Because the parcel is saturated, its mixing ratio must equal the saturation mixing ratio at the freezing level. The isopleth of saturation mixing ratio passing through state 3 gives for the mixing ratio there

$$r_{FL} = 5.5 \text{ g kg}^{-1}.$$

If no precipitation occurs, the total water content of the parcel is preserved. Therefore, the liquid water content at the freezing level is given by

$$r_l = 0.5 \text{ g kg}^{-1}.$$

Approximately 10% of the parcel's moisture has condensed by this altitude.

Temperature inside cloud at 650 hPa

At 650 hPa, the parcel has evolved along the $\theta_e = 315$ K pseudo adiabat to state 4, where its temperature is

$$T_{650} = -6 \text{ C.}$$

Were the parcel perfectly dry, it would have continued to evolve above the LCL (770 hPa) along the 315 K adiabat. Without the release of latent heat to offset adiabatic cooling, the parcel's temperature would then have decreased more rapidly, resulting in a temperature at 650 hPa of -12°C .

Mixing ratio inside cloud at 650 hPa

Because the parcel remains saturated, the isopleth of saturation mixing ratio that passes through state 4 gives for the mixing ratio at 650 hPa

$$r_{650} = 4.0 \text{ g kg}^{-1}.$$

By this level, one third of the parcel's water vapor has been transformed into condensate.

This example illustrates the strong constraint on water vapor imposed by thermodynamics and hydrostatic stratification. Inside convective tower, which transports moisture upward from the Earth's surface, the abundance of vapor that can be supported in solution with dry air decreases with altitude (Fig. 5.3). By 500 hPa, less than 30% of the surface mixing ratio of water inside the parcel under consideration remains as vapor. According to (1.24), the absolute humidity ρ_v decreases with altitude even faster. Less than 15% of the absolute concentration of water vapor remains after

⁷ Ice forms only in the presence of a special type of aerosol particle, referred to as a *freezing nucleus* (Chap. 9). Because freezing nuclei are comparatively rare, many cloud droplets are actually "supercooled": They remain liquid at temperatures below 0°C .

the parcel has been displaced to the middle troposphere. It is for this reason that deep convection increases the column abundance of water vapor in Fig. 1.19 only modestly. Much of Σ_{H_2O} resides in the lowest 1–2 km, where water vapor is distributed in the horizontal more uniformly than it is at higher altitude (cf. Fig. 1.18).

Through this process, water vapor that is transported upward inside convective tower is systematically removed. Condensation leads to a decrease with altitude of mixing ratio and, via precipitation, a similar decrease of total water content. Outside convective tower, air is even drier – because it has not recently been in contact with the source of water vapor at the Earth's surface or because it has been dehydrated inside convective tower. By restricting vertical transport of water, thermodynamics, together with hydrostatic stratification, confines moisture to a shallow neighborhood of the Earth's surface. It thereby isolates water from photo-dissociating radiation at higher altitude, a process that has preserved water on the Earth.

SUGGESTED REFERENCES

A complete discussion of the Gibbs-Dalton law, along with its implications to a multi-component system, is given in *Thermodynamics* (1970) by J. Keenan. *Atmospheric Thermodynamics* (1981) by Iribarne and Godson contains a detailed treatment of moisture-dependent properties and a survey of thermodynamic charts. *The Principles of Chemical Equilibrium* (1971) by K. Denbigh discusses the thermodynamics of condensed phases.

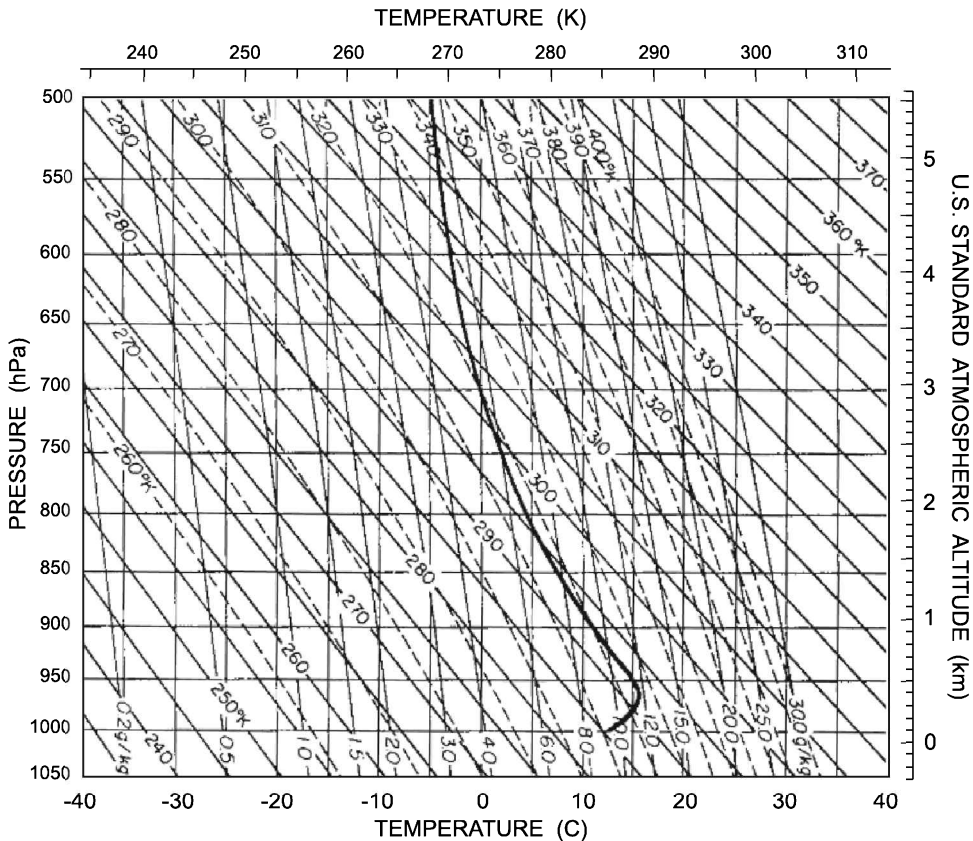
PROBLEMS

The pseudo-adiabatic chart in Appendix F is to be used only for those problems in which it is explicitly indicated.

1. Downslope winds in North America are called a *chinook*, an Indian term for “snow eater.” A chinook often occurs with the mountains blanketed in cloud but with clear skies leeward. Consider the following synoptic situation: Moist air originating in the eastern Pacific is advected over the western United States, where it is forced over the continental divide. On the windward side, the surface lies at 800 hPa, where the temperature and mixing ratio are 20°C and 15 g kg⁻¹, respectively. If the summit lies at 600 hPa and if any condensate that forms precipitates out, determine (a) the surface air temperature at Denver, which lies at 830 hPa, (b) the mixing ratio at Denver, (c) the relative humidity at Denver.
2. A refrigerator having an interior volume of 2.0 m³ is sealed and switched on. If the air initially has a temperature of 30°C and a relative humidity of 0.50, determine (a) the temperature at which condensation forms on the walls, (b) how much moisture will have condensed when the temperature reaches 2°C, (c) how much heat must be rejected to the surroundings to achieve the final state in (b).
3. State the number of thermodynamic degrees of freedom for (a) moist unsaturated air, (b) moist saturated air. (c) Explain the numbers in (a) and (b) and why they differ.
4. Compare θ and θ_v for saturated conditions at 1000 hPa and a temperature of (a) 10°C, (b) 20°C, (c) 30°C.
5. Warm moist air leaves an array of cooling towers at a power plant situated at 825 hPa. A cloud forms directly overhead. If the ambient temperature profile is isothermal, with $T = 5^\circ\text{C}$, and if the initial temperature and mixing ratio of air leaving the towers are 30°C and 25 g kg⁻¹, respectively, determine (a) the relative

humidity immediately above the towers, (b) the dew point immediately above the towers, (c) the virtual potential temperature immediately above the towers; compare this value with the potential temperature and discuss it in relation to typical differences between θ and θ_v , (d) the pressure at cloud base, (e) the equivalent potential temperature at 800 hPa, (f) the mixing ratio at 700 hPa, (g) the pressure at cloud top. A pseudo-adiabatic chart is provided in Appendix F.

6. Room temperature on a given day is 22°C, whereas the outside temperature at 1000 hPa is 2°C. Calculate the maximum relative humidity that can be accommodated inside without room windows fogging, if the windows can be treated as having a uniform temperature.
7. (a) Under the same conditions as in (6), but for an aircraft cabin that is pressurized to 800 hPa. (b) As in (a), but noting that the interior pane is thermally isolated from the exterior pane.
8. Evaporation is an efficient means of cooling air and, through contact with other media, of rejecting heat to the environment. An evaporative cooler processes ambient air at 35°C and 900 hPa to produce room air with a relative humidity of 0.60. Calculate the temperature inside if the relative humidity outside is (a) 0.10, (b) 0.50.
9. A cooling tower situated at 1000 hPa processes ambient air to release saturated air at a temperature of 35°C and at a rate of 10 m³s⁻¹. Calculate the rate heat is rejected if ambient air has a temperature of 20°C and a relative humidity of (a) 0.20, (b) 0.80.



10. A morning temperature sounding is plotted above. Through absorption of SW radiation at the ground, the surface inversion ($\Gamma < 0$) that developed during the night is replaced during the day by adiabatic thermal structure in the lowest half kilometer. A cumulus cloud then forms over an asphalt parking lot at 1000 hPa, where the mixing ratio is 10 g kg^{-1} and the relative humidity is 50%. For the air column above the parking lot, determine the (a) surface air temperature, (b) pressure at the cloud base, (c) potential temperature at 900 hPa, (d) potential temperature at 700 hPa, (e) equivalent potential temperature at the surface, (f) mixing ratio at 700 hPa, (g) pressure at the cloud top, (i) mixing ratio at the cloud top. A pseudo-adiabatic chart is provided in Appendix F.
11. Moist air moves inland from a maritime region, where it has a temperature of 16°C and a relative humidity of 66%. Through contact with the ground, the air cools. Calculate the temperature at which fog forms.
12. Use the pseudo-adiabatic chart in Appendix F to determine the LCL above terrain at 1000 hPa for (a) moist surface conditions representative of the eastern US: $T = 30^\circ\text{C}$ and $RH = 70\%$, (b) arid surface conditions representative of the southwestern US: $T = 30^\circ\text{C}$ and $RH = 10\%$. (c) Contrast the heights of cumulus cloud bases under these conditions in relation to the likelihood of precipitation reaching the surface.
13. An altitude chamber is used to simulate a sudden decompression from normal cabin pressure: 800 hPa, to ambient pressure at 18,000'. If the initial temperature is 22°C , how small must the relative humidity be to avoid spontaneous cloud formation during the decompression? A pseudo-adiabatic chart is provided in Appendix F.
14. Outside air has a temperature of -10°C and a relative humidity of 0.50. (a) What is the relative humidity indoors if the room temperature is 22°C and if air is simply heated, without humidification? (b) What mass of water vapor must be added to a room volume of 75 m^3 to elevate its relative humidity to 40%? (c) How much energy is required to achieve this state?
15. On a given day, the lapse rate and relative humidity are constant, with $\Gamma = 8^\circ\text{K km}^{-1}$ and $RH = 0.80$. If the surface temperature at 1000 hPa is 20°C , (a) estimate the total precipitable water vapor. (b) Below what height is 90% of the water vapor column represented? (c) Now calculate the precipitable water vapor inside a column of ascending air in a cumulonimbus tower that extends to 100 hPa. A pseudo-adiabatic chart is provided in Appendix F.
16. A morning temperature sounding over Florida reveals the profile

$$T = \begin{cases} 10 - 6(z - 1) (\text{C}) & z \geq 1 \text{ km} \\ 20 - 10z (\text{C}) & z < 1 \text{ km.} \end{cases}$$

That afternoon, surface conditions are characterized by a temperature of 30°C and a mixing ratio of 20 g kg^{-1} . (a) To what height will cumulus clouds develop if air motion is nearly adiabatic? (b) Plot the vertical profile of mixing ratio from the surface to the top of a cumulus cloud. (c) Over the same altitude range, plot the vertical profile of mixing ratio but neglecting the explicit pressure dependence of saturation mixing ratio. (d) Discuss how the results in (a) and (b) would be modified if air inside the ascending plume is not adiabatic but, rather, mixes with surrounding air in roughly equal proportion. A pseudo-adiabatic chart is provided in Appendix F.

17. For the temperature sounding and surface conditions in (16), determine the height of convection and the vertical profile of temperature inside an ascending plume were the air perfectly dry. A pseudo-adiabatic chart is provided in Appendix F.
18. Graph the dew point temperature as a function of relative humidity for 1000 hPa and 20°C.
19. For reasons developed in Chap. 9, many clouds are supercooled (e.g., they contain liquid droplets at temperatures below 0°C). For cloud that is saturated with respect to water, calculate the relative humidity with respect to ice at a temperature of (a) -5°C, (b) -20°C.
20. Inspection of the pseudo-adiabatic chart reveals that, as $z \rightarrow \infty$, pseudo-adiabats ($\theta_e = \text{const}$) approach adiabats ($\theta = \text{const}$). Why?

21. The development in Sec. 5.1 for a mixture of dry air and water considers the presence of only one condensed phase. Why and under what conditions is this permissible?

22. A cold front described by the surface

$$z = h(x - ct)$$

moves eastward with velocity c and undercuts warmer air ahead of it. Far ahead of the front, undisturbed air is characterized by the temperature and mixing ratio profiles $T_\infty(z)$ and $r_\infty(z)$, respectively. If air is lifted in unison over the frontal surface, for a fixed location x , derive an expression for (a) the variation of temperature with altitude above the frontal surface but beneath the LCL as a function of time, (b) the variation of mixing ratio with altitude above the frontal surface but beneath the LCL as a function of time, (c) an expression for the altitude of the LCL if, far from the frontal zone, pressure decreases with altitude approximately as $e^{-\frac{z}{H}}$, with $H = \text{const}$.

23. Show that the exponent in (5.26) vanishes for an air parcel that is displaced to the top of the atmosphere.
24. Revisit Prob. 2.14, recognizing now that the bubble is surrounded by water. If the bubble is saturated throughout its ascent, yet remains adiabatic, calculate the bubble's temperature on reaching the surface.
25. An air parcel drawn into a cumulonimbus tower becomes saturated at 900 hPa, where its temperature is 20°C. Calculate the fractional internal energy represented by condensate when the parcel has been displaced to 100 hPa. A pseudo-adiabatic chart is provided in Appendix F.
26. The base of convective cloud may be estimated from the dew point spread at the surface. (a) Derive an expression for the dew point temperature T_d of an ascending parcel as a function of its height in terms of the surface mixing ratio r_0 , presuming environmental conditions to be isothermal. (b) Show that T_d for the parcel varies with height approximately linearly in the lowest few kms. (c) Use the lapse rate of T_d in (b) to derive an expression for the height of the LCL in terms of the dew point spread at the surface: $(T - T_d)_0$. (d) Use the result of (c) to estimate the cloud base under the conditions in (16).

Hydrostatic equilibrium

Changes of thermodynamic state that accompany vertical motion follow from the distribution of atmospheric mass, which is determined ultimately by gravity. In the absence of motion, Newton's second law applied to the vertical reduces to a statement of *hydrostatic equilibrium* (1.16). Gravity is then balanced by the vertical pressure-gradient force. This simple form of mechanical equilibrium is accurate even in the presence of motion because the acceleration of gravity is, almost invariably, much greater than vertical acceleration of individual air parcels. Only inside deep convective towers and other small-scale phenomena is vertical acceleration large enough to invalidate hydrostatic equilibrium.

Because it is such a strong body force, gravity must be treated with some care. Complications arise from the fact that the gravitational acceleration experienced by an air parcel does not act purely in the vertical. It also varies with location. According to the preceding discussion, gravity is large enough to overwhelm other contributions in the balance of vertical forces. The same holds for the balance of horizontal forces. Horizontal components of gravity that are introduced by the Earth's rotation and other sources must be balanced by additional horizontal forces. Unrelated to air motion, those additional forces unnecessarily complicate the description of atmospheric motion.

6.1 EFFECTIVE GRAVITY

In the reference frame of the Earth, the gravitational acceleration experienced by an air parcel involves three basic contributions:

- (1) radial gravitation by the Earth's mass,
- (2) centrifugal acceleration due to rotation of the reference frame, and
- (3) anisotropic contributions,

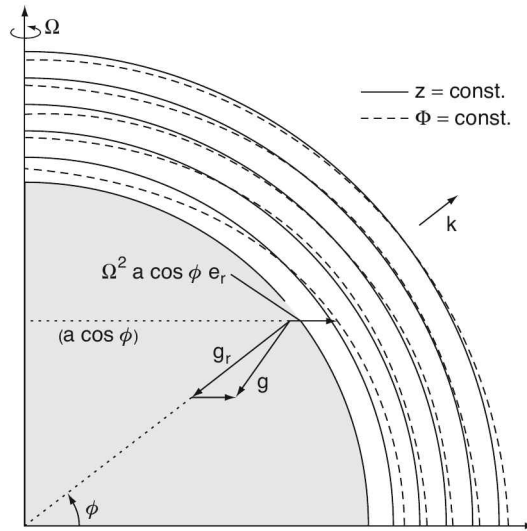


Figure 6.1 Effective gravity \mathbf{g} illustrated in relation to its contributions from radial gravitation and centrifugal acceleration. Also shown are surfaces of constant geometric altitude z and corresponding surfaces of constant geopotential Φ , to which \mathbf{g} is orthogonal.

as illustrated in Fig. 6.1. Radial gravitation by the Earth's mass is the dominant contribution. Under idealized circumstances, that component of \mathbf{g} acts perpendicular to surfaces of constant geometric altitude z , which are concentric spheres. The Earth's rotation introduces another contribution. Because it is noninertial, the reference frame of the Earth includes a centrifugal acceleration. It acts perpendicular to and away from the axis of rotation. Lastly, departures of the Earth from sphericity and homogeneity (e.g., through variations of surface topography) introduce anisotropic contributions to gravity. They must be determined empirically.

Collectively, these contributions determine the *effective gravity*:

$$\mathbf{g}(\lambda, \phi, z) = \frac{a^2}{(a+z)^2} g_0 \mathbf{k} + \Omega^2 (a+z) \cos \phi \mathbf{e}_r + \boldsymbol{\epsilon}(\lambda, \phi, z), \quad (6.1)$$

where a is the mean radius of the Earth, g_0 is the radial gravitation at mean sea level, \mathbf{k} is the local upward unit normal, Ω is the Earth's angular velocity, λ and ϕ are longitude and latitude, respectively, \mathbf{e}_r is a unit vector directed outward from the axis of rotation, and $\boldsymbol{\epsilon}(\lambda, \phi, z)$ represents all anisotropic contributions. Even if the Earth were perfectly spherical, effective gravity would not act uniformly across the surface of the Earth. Nor would it act entirely in the vertical. According to (6.1), the centrifugal acceleration deflects \mathbf{g} into the horizontal, introducing a component of gravity along surfaces of constant z . The centrifugal acceleration also causes the vertical component of \mathbf{g} to vary with location, from a minimum of 9.78 m s^{-2} at the equator to a maximum of 9.83 m s^{-2} at the poles. Anisotropic contributions, although more complicated, have a similar effect.

These contributions introduce horizontal components of gravity. They must be balanced by other forces that act along surfaces of constant geometric altitude. Thus,

(1.16) represents but one of several component equations that are required to describe hydrostatic equilibrium. Horizontal forces that are introduced by gravity can overshadow those which actually control the motion of an air parcel. For instance, rotation introduces a horizontal component that must be balanced by a variation of pressure along surfaces of constant geometric altitude. That horizontal pressure gradient exists even under static conditions, wherein an observer in the reference frame of the Earth would anticipate Newton's second law to reduce to simple hydrostatic balance in the vertical. To avoid such complications, it is convenient to introduce a coordinate system that consolidates all components of gravity into a single vertical component.

6.2 GEOPOTENTIAL COORDINATES

Gravity is conservative. The specific work performed during a cyclic displacement of mass through the Earth's gravitational field therefore vanishes

$$\Delta w_g = - \oint \mathbf{g}(\lambda, \phi, z) \cdot d\boldsymbol{\xi} = 0, \quad (6.2)$$

where $d\boldsymbol{\xi}$ denotes the incremental displacement of an air parcel. According to Sec. 2.1.4, (6.2) is a necessary and sufficient condition for the existence of an exact differential

$$\begin{aligned} d\Phi &= \delta w_g \\ &= -\mathbf{g} \cdot d\boldsymbol{\xi}, \end{aligned} \quad (6.3)$$

which defines the gravitational potential or *geopotential* Φ . In (6.3), $d\Phi$ equals the specific work performed against gravity to complete the displacement $d\boldsymbol{\xi}$. Then, per discussion in Sec. 2.1.4, gravity is an irrotational vector field that may be expressed in terms of its potential function

$$\mathbf{g} = -\nabla\Phi. \quad (6.4)$$

Surfaces of constant geopotential are not spherical like surfaces of constant geometric altitude (Fig. 6.1). Centrifugal acceleration in the rotating reference frame of the Earth distorts geopotential surfaces into oblate spheroids. The component of $\nabla\Phi$ along surfaces of constant geometric altitude introduces a horizontal component of gravity. It must be balanced by a horizontal pressure gradient force.

Representing gravity is simplified by transforming from pure spherical coordinates, in which elevation is fixed along surfaces of constant geometric altitude, to geopotential coordinates, in which elevation is fixed along surfaces of constant geopotential. Geopotential surfaces may be used for coordinate surfaces because Φ increases monotonically with altitude (6.3), ensuring a one-to-one relationship between those variables. Introducing the aforementioned transformation is tantamount to measuring elevation z along the line of effective gravity. The latter will be termed *height* to distinguish it from geometric altitude. Defining z in this manner agrees with the usual notion of local vertical being plumb with the line of gravity. It will be adopted as the convention hereafter.

In geopotential coordinates, gravity has no horizontal component. Consequently, in terms of height, (6.4) reduces to simply

$$d\Phi = g dz, \quad (6.5)$$

where g denotes the magnitude of \mathbf{g} . The latter acts in the direction of decreasing height - not geometric altitude. Using mean sea level as a reference elevation¹ yields an expression for the absolute geopotential

$$\Phi(z) = \int_0^z g dz'. \quad (6.6)$$

Having dimensions of specific energy, the geopotential represents the work that must be performed against gravity to raise a unit mass from mean sea level to a height z . Evaluating Φ from (6.6) requires measured values of effective gravity throughout. However, because it is a point function, Φ is independent of path by (6.2) and the Exact Differential Theorem (2.7). Therefore, Φ depends only on height z . For the same reason, surfaces of constant geopotential coincide with surfaces of constant height.

Using surfaces of constant geopotential for coordinate surfaces consolidates gravity into the vertical coordinate, height. However, g still varies with z through (6.5). To account for that variation, it is convenient to introduce yet another vertical coordinate in which the dependence on height is absorbed. *Geopotential height* is defined as

$$\begin{aligned} Z &= \frac{1}{g_0} \int_0^z g dz \\ &= \frac{1}{g_0} \Phi(z), \end{aligned} \quad (6.7)$$

where $g_0 = 9.8 \text{ m}^2 \text{ s}^{-1}$ is a reference value reflecting the average over the surface of the Earth. Then

$$\begin{aligned} dZ &= \frac{1}{g_0} d\Phi \\ &= \frac{g}{g_0} dz. \end{aligned} \quad (6.8)$$

Because it accounts for the variation of gravity, geopotential height simplifies the expression for hydrostatic balance. In terms of Z , (1.16) reduces to

$$\begin{aligned} dp &= -\rho g dz \\ &= -\rho g_0 dZ. \end{aligned} \quad (6.9)$$

When elevation is represented in terms of geopotential height, gravity is described by the constant value g_0 .

While having these formal advantages, geopotential height is nearly identical to height because $g \cong g_0$ throughout the homosphere. In fact, the two measures of elevation differ by less than 1% below 60 km. Therefore, Z may be used interchangeably with z . This will be our convention hereafter. Then g is understood to refer to the constant reference value g_0 .

6.3 HYDROSTATIC BALANCE

In terms of geopotential height, hydrostatic balance is expressed by

$$g dz = -v dp. \quad (6.10)$$

¹ Mean sea level is nearly coincident with a surface of constant geopotential.

The ideal gas law transforms this into

$$\begin{aligned} dz &= -\frac{RT}{g} d\ln p \\ &= -H d\ln p, \end{aligned} \quad (6.11)$$

where the scale height H is defined in (1.17). According to (6.11), the change of height is proportional to the local temperature and to the change of $-\ln p$. As vertical changes of temperature and H are small compared with changes of pressure, the quantity $-\ln p$ serves as a dimensionless measure of height.

6.3.1 Hypsometric equation

Consider a layer bounded by two isobaric surfaces $p = p_1(x, y, z)$ and $p = p_2(x, y, z)$, as illustrated in Fig. 6.2. Integrating (6.11) between those surfaces obtains

$$\Delta z = z_2 - z_1 = -\langle H \rangle \ln \left(\frac{p_2}{p_1} \right), \quad (6.12.1)$$

where

$$\langle H \rangle = \frac{R}{g} \langle T \rangle \quad (6.12.2)$$

and

$$\langle T \rangle = \frac{\int_{p_1}^{p_2} T d\ln p}{\int_{p_1}^{p_2} d\ln p} \quad (6.12.3)$$

define the layer-mean scale height and temperature, respectively, and $z_1 = z(x, y, p_1)$ and $z_2 = z(x, y, p_2)$ denote the heights of the bounding isobaric surfaces. Thus

$$z_2 - z_1 = \frac{R}{g} \int_{p_2}^{p_1} T d\ln p. \quad (6.12.4)$$

Known as the *hypsometric equation*, (6.12) asserts that the *thickness* of a layer bounded by two isobaric surfaces is proportional to the mean temperature of that layer and the pressure difference across it.

In regions of cold air, the e-folding scale H is short. Pressure then decreases upward sharply (1.17). The vertical separation of isobaric surfaces in such regions is therefore compressed (Fig. 6.2). In regions of warm air, H is tall. Pressure then decreases upward slowly. The separation of isobaric surfaces in those regions is therefore expanded.

6.3.2 Meteorological Analyses

If $z_1 = 0$, the hypsometric equation provides the height of the upper surface $z_2 = z(x, y, p_2)$. For example, the height of the 500-hPa surface in Fig. 1.9 follows as a vertical integral of temperature. It may be determined from measurements of temperature between that surface and mean sea level (~ 1000 hPa) or, alternatively, between that surface and a reference isobaric surface, if the height of the reference surface is known. Such temperature measurements are made routinely by rawinsondes (Sec. 1.6.1). Also measuring air motion and humidity, rawinsondes ascend via balloon twice-daily. They provide coverage that is dense over populated continent, but sparse over ocean and

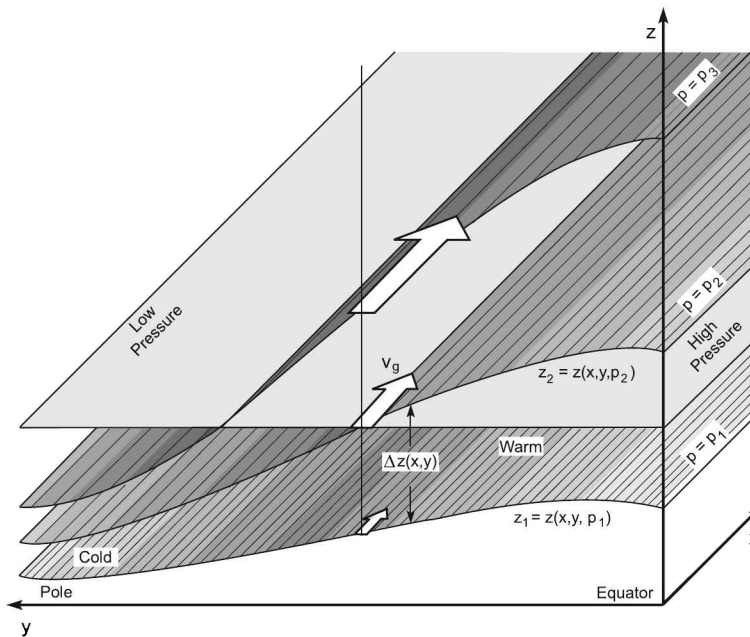


Figure 6.2 Isobaric surfaces (stippled) in the presence of a horizontal temperature gradient. Vertical spacing of isobaric surfaces is compressed in cold air and expanded in warm air, which makes the height of an individual isobaric surface low in the former and high in the latter. Because pressure decreases upward monotonically, a surface of constant height (shaded) then has low pressure where isobaric height is low and high pressure where isobaric height is high. Therefore, contours of height for an individual isobaric surface may be interpreted similarly to isobars on a constant height surface. Also indicated is the geostrophic velocity v_g (Sec. 12.1), which is directed parallel to contours of isobaric height and has magnitude proportional to its gradient.

in the tropics (Fig. 1.37). Temperature measurements are also made remotely by satellite. Satellite measurements of temperature provide continuous global coverage, albeit *asynoptic*.² Satellite retrievals also have coarser vertical resolution ($\sim 4\text{--}8$ km) than that provided by rawinsondes (~ 100 m). Both are assimilated operationally in *synoptic analyses* of the instantaneous circulation, like those presented in Figs. 1.9 and 1.10.

Synoptic analyses fall into two categories:

- (1) The *surface analysis* displays properties at the ground, but extrapolated to a single reference level: mean sea level. It thus displays sea level pressure (SLP) on a surface of constant height $z = 0$.

² The term *synoptic* refers to the simultaneous distribution of some property, for example, a snapshot over the Earth of the instantaneous height of the 500-hPa surface (Fig. 1.9a). A polar-orbiting satellite, however, observes different sites at different times. Termed *asynoptic*, such measurements sample the entire Earth only after 12–24 hours.

- (2) An *upper air analysis* displays properties on a surface of constant pressure. Accordingly, an upper air analysis is an *isobaric analysis*: It represents a quasi-horizontal section of the circulation along a surface where $p = \text{const.}$ ³

Figure 1.9 displays the 500-hPa analysis. On an isobaric surface, displaying pressure is uninformative; it's constant. Displayed instead is the height of the 500-hPa surface. This representation displays height on a surface of constant p . It thus interchanges the roles of p and z . Elevation is then represented by pressure, which becomes the vertical coordinate (e.g., $p = 700$ hPa, 500 hPa, 300 hPa, etc).

In isobaric coordinates, pressure becomes the independent variable. The height of an isobaric surface then becomes the dependent variable, for example, $z_{500} = z(x, y, 500 \text{ hPa})$. A theoretical basis for using pressure to describe elevation follows from the hypsometric equation. Because H is positive, (6.12) implies a one-to-one relationship between height and pressure: A particular height corresponds to a particular pressure. Pressure may therefore be used alternatively to height as the vertical coordinate. Isobaric coordinates, developed in Chap. 11, use surfaces of constant pressure for coordinate surfaces. Although they evolve with the circulation, isobaric coordinates afford several advantages.

The horizontal distribution of height on an isobaric surface, $z = z(x, y, p)$, may be interpreted analogously to the distribution of pressure on a surface of constant height, $p = p(x, y, z)$, for example, on the mean height of that isobaric surface. Pressure decreases with height monotonically. Consequently, low height of an isobaric surface corresponds to low pressure on a surface of constant height (Fig. 6.2). Just the reverse applies to high values. For this reason, contours of height on an isobaric surface mirror contours of pressure on a surface of constant height. From hydrostatic equilibrium, that pressure represents the weight of the overlying atmospheric column. The distribution of isobaric height therefore reflects the horizontal distribution of atmospheric mass.

According to Fig. 1.9, the height of the 500-hPa surface slopes downward toward the pole at mid-latitudes, where z_{500} decreases from 5900 m to 5000 m. The decrease of 500-hPa height mirrors the poleward decrease of layer-mean temperature between the surface and 500 hPa. $\langle T \rangle_{500}$ is plotted in Fig. 6.3 for the same time as in Fig. 1.9a. The wavy region where $\langle T \rangle_{500}$ and z_{500} decrease sharply delineates the *polar front*. It separates cold polar air from warmer air at lower latitude. As is apparent from Fig. 1.9a, that region of sharp temperature gradient coincides with strong westerlies of the jet stream, which are tangential to contours of isobaric height. The jet stream thus marks the interface between cold polar air and warmer air at lower latitude. A steepening of the meridional temperature gradient steepens the gradient of 500-hPa height (6.12). That, in turn, contracts the horizontal spacing of height contours in Fig. 1.9a. For reasons developed in Chap. 12, this contraction is accompanied by an intensification of the jet.

Meridional gradients of temperature and height are particularly steep east of Asia and North America. Those features coincide with the North Pacific and North Atlantic storm tracks. They are manifest in the time-mean circulation (Fig. 1.9b) as local

³ This representation is motivated by observational considerations. Sounding measurements at different sites are returned at specified values of pressure, which are interpolated seamlessly along an isobaric surface.

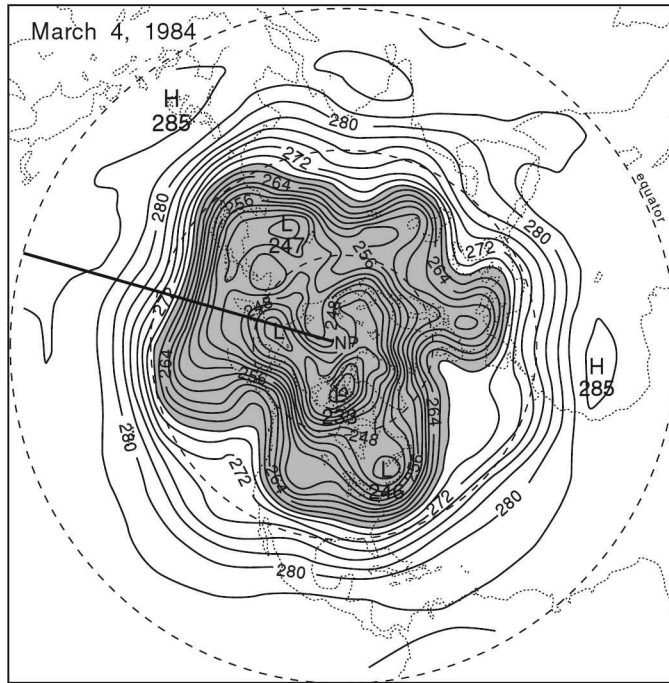


Figure 6.3 Mean temperature ($^{\circ}\text{K}$) of the layer between 1000 hPa and 500 hPa for March 4, 1984. Bold solid line marks a meridional cross section displayed in Fig. 6.4.

intensifications of the jet stream. Anomalies of cold air that punctuate the polar front in Fig. 6.3 (e.g., the cold outbreak over the North Atlantic) introduce synoptic-scale depressions of the 500-hPa surface in Fig. 1.9a. By deflecting contours of isobaric height, those features disrupt the jet stream. In their absence, the jet would be nearly zonal (e.g., Fig. 1.9b).

Contours of 500-hPa height that delineate the polar front in Fig. 1.9a correspond to isobars on the constant-height surface: $z \sim 5.5$ km, with pressure decreasing toward the pole. Thus less atmospheric column lies above that elevation poleward of the front than equatorward of the front. Figure 6.4 presents a meridional cross section of thermal structure and motion in Figs. 6.3 and 1.9a. The horizontal distribution of mass just noted results from compression of isobaric surfaces poleward of the front, which is positioned near 40 N, and expansion of those surfaces equatorward of the front. This introduces a tilt to isobaric surfaces at midlatitudes, one which steepens with height. Found at the same latitudes is the jet stream. Like the tilt of isobaric surfaces, westerlies intensify upward, exceeding 75 m/s near the tropopause. Similar behavior is evident at high latitude. It marks the base of the polar-night jet, which intensifies in the stratosphere.

6.4 STRATIFICATION

The lapse rates Γ_d and Γ_s , (2.33) and (5.32), describe the evolution of a displaced air parcel under unsaturated and saturated conditions, respectively. They apply formally

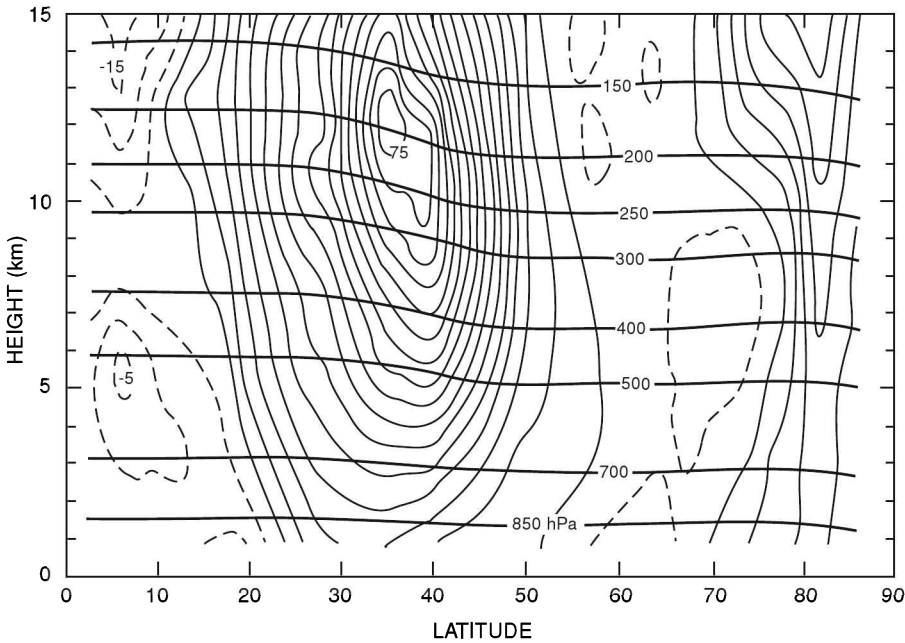


Figure 6.4 Meridional cross section of thermal structure and motion in Figs. 1.9a and 6.3 at the position indicated in the latter. Vertical spacing of isobaric surfaces (heavy solid curves) is expanded at low latitude and compressed at high latitude, introducing a slope at mid-latitudes that steepens with height. Zonal wind speed (contoured in m s^{-1}) delineates the subtropical jet, which coincides with the steep slope of isobaric surfaces. Both intensify upward to a maximum at the tropopause. Contour increment: 5 m s^{-1} .

with z representing geopotential height and g constant. Neither, however, has a direct relationship to the temperature of the parcel's environment – because, under adiabatic conditions, a displaced parcel is thermally isolated.

Thermal properties of the environment are dictated by the history of air at a given location, namely, by where that air has been and what thermodynamic influences it has experienced. The *environmental lapse rate* is defined as

$$\Gamma = -\frac{dT}{dz}, \quad (6.13)$$

where T refers to the ambient temperature. Like parcel lapse rates, $\Gamma > 0$ corresponds to temperature decreasing with height. Conversely, $\Gamma < 0$ corresponds to temperature increasing with height. In the latter case, the profile of environmental temperature is said to be *inverted*.

The compressibility of air leads to atmospheric mass being stratified, as is reflected in the sharp upward decrease of density and pressure (Fig. 1.2). Because the environment is in hydrostatic equilibrium, the distribution of pressure is related to thermal structure through the hypsometric relation. Incorporating (6.11) transforms (6.13) into

$$-\frac{1}{H} \frac{dT}{d \ln p} = \Gamma$$

or

$$\begin{aligned}\frac{d \ln T}{d \ln p} &= \frac{R}{g} \Gamma \\ &= \kappa \frac{\Gamma}{\Gamma_d}.\end{aligned}\quad (6.14)$$

6.4.1 Idealized stratification

For certain classes of thermal structure, (6.14) can be integrated analytically to obtain the distributions of pressure and other thermal properties inside an atmospheric layer.

Layer of constant lapse rate

If $\Gamma = \text{const} \neq 0$, (6.14) yields

$$\frac{T}{T_s} = \left(\frac{p}{p_s} \right)^{\frac{\Gamma}{\Gamma_d} \kappa}, \quad (6.15)$$

where the subscript s refers to the base of the layer. Because

$$T(z) = T_s - \Gamma z, \quad (6.16)$$

(6.15) may be used to obtain the vertical distribution of pressure

$$\frac{p}{p_s} = \left[1 - \kappa \frac{\Gamma}{\Gamma_d} \left(\frac{z}{H_s} \right) \right]^{\frac{\Gamma_d}{\kappa \Gamma}}. \quad (6.17)$$

Then (2.31) implies the vertical profile of potential temperature

$$\theta(z) = (T_s - \Gamma z) \left[1 - \kappa \frac{\Gamma}{\Gamma_d} \left(\frac{z}{H_s} \right) \right]^{-\frac{\Gamma_d}{\Gamma}}, \quad (6.18)$$

where $p_s = p_0 = 1000$ hPa has been presumed for the base of the layer.

As shown in Fig. 6.5, the compressibility of air makes pressure decrease with height for all Γ . The same is true of density. However, potential temperature decreases with height only for $\Gamma > \Gamma_d$. It increases with height for $\Gamma < \Gamma_d$. In both cases, θ varies sharply with height because the pressure term in (6.18) dominates over the temperature term. As a result, environmental potential temperature can change by several hundred Kelvin in just a couple scale heights. Such thermal structure is typical in the stratosphere (cf. Fig. 18.19). In contrast, the potential temperature of a displaced air parcel is independent of height, at least under unsaturated conditions.

Thermal structure described by (6.16)–(6.18) can apply locally (e.g., inside a layer), even if temperature does not vary linearly throughout the atmosphere. Should it apply for the entire range of height, (6.16) has one further implication. For $\Gamma > 0$, positive temperature requires the atmosphere to have a finite upper bound

$$z_{\text{top}} = \frac{T_s}{\Gamma}, \quad (6.19)$$

where p vanishes. For $\Gamma \leq 0$, no upper bound exists.

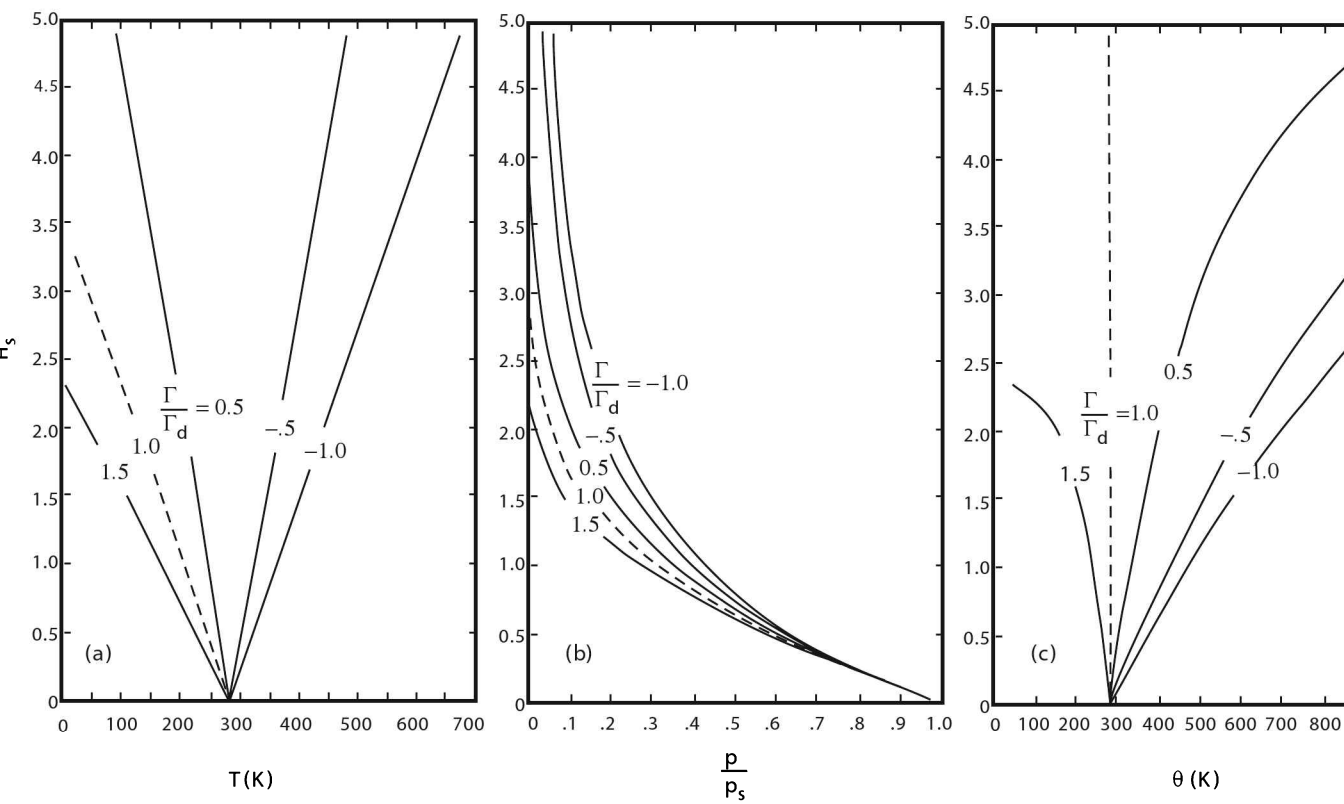


Figure 6.5 Vertical profiles of (a) temperature, (b) pressure, and (c) potential temperature inside layers of constant lapse rate Γ .

Isothermal layer

For the special case $\Gamma = 0$, (6.17) is indeterminate. Reverting to the hydrostatic relation (6.11) yields

$$z = -H \ln \left(\frac{p}{p_s} \right)$$

or

$$\frac{p}{p_s} = e^{-\frac{z}{H}}, \quad (6.20)$$

which is identical to (1.17) under these circumstances. Then the vertical profile of potential temperature is simply

$$\begin{aligned} \theta(z) &= T_s e^{\kappa \frac{z}{H}} \\ &= \theta_s e^{\kappa \frac{z}{H}}. \end{aligned} \quad (6.21)$$

Under isothermal conditions, environmental potential temperature increases by a factor of e every $\kappa^{-1} \cong 3$ scale heights.

Adiabatic layer

For the special case $\Gamma = \Gamma_d$, (6.15) reduces to

$$\frac{T}{T_s} = \left(\frac{p}{p_s} \right)^\kappa. \quad (6.22)$$

The distribution of pressure (6.17) then becomes

$$\frac{p}{p_s} = \left[1 - \kappa \frac{z}{H_s} \right]^{\frac{1}{\kappa}}. \quad (6.23)$$

Equation (6.22) is the same relationship between temperature and pressure implied by Poisson's equation (2.30.2) that defines potential temperature (2.31). Accordingly, the vertical profile of potential temperature (6.18) reduces to

$$\theta(z) = \theta_s = \text{const.} \quad (6.24)$$

6.5 LAGRANGIAN INTERPRETATION OF STRATIFICATION

As noted earlier, hydrostatic equilibrium applies in the presence of motion as well as under static conditions. Therefore, each of the stratifications in Sec. 6.4 is valid even if a circulation is present, as is invariably the case. Under those circumstances, vertical profiles of temperature, pressure, and potential temperature (6.16)–(6.18) correspond to horizontal-mean thermal structure. They may be interpreted as averages over many ascending and descending air parcels. Interpreting thermal structure in this manner provides some insight into the mechanisms that shape mean stratification.

For a layer of constant lapse rate, the relationship between temperature and pressure (6.15) resembles one implied by Poisson's equation (2.30.2), but for a polytropic process (Sec. 2.5.1) with $\frac{\Gamma}{\Gamma_d} \kappa$ in place of κ . Consequently, we may associate the thermal structure in (6.16)–(6.18) with a vertical rearrangement of air in which individual parcels evolve diabatically according to a polytropic process. In that

description, air parcels moving vertically exchange heat with their surroundings in such proportion for their temperatures to vary linearly with height. For an individual air parcel, the foregoing process has a polytropic specific heat that satisfies

$$\frac{R}{(c_p - c)} = \frac{\Gamma}{\Gamma_d} \kappa$$

or

$$c = c_p \left(1 - \frac{\Gamma_d}{\Gamma} \right). \quad (6.25)$$

The corresponding heat transfer for the parcel is then described by (2.37). Its change of potential temperature follows from (2.40).

6.5.1 Adiabatic stratification: A paradigm of the troposphere

Consider a layer characterized by

$$\Gamma = \Gamma_d. \quad (6.26.1)$$

Such stratification is termed *adiabatic*. Then (6.25) implies

$$\begin{aligned} c &= 0 \\ \delta q &= d\theta = 0. \end{aligned} \quad (6.26.2)$$

Thus, individual air parcels evolve adiabatically. Under these circumstances, air parcels undergo vertical motion but do not interact thermally with their surroundings. Such behavior may be regarded as the limiting case of when vertical motion is fast compared with diabatic influences. It is typical of cumulus and sloping convection. Although the vertical motion of parcels is adiabatic, they may still exchange heat at the boundaries of the layer. In fact, in the absence of mechanical forcing, such heat transfer is necessary to maintain vertical motion.

The evolution of an individual parcel may be treated as a circuit that it follows between the upper and lower boundaries of the layer, where heat transfer is concentrated. Such a circuit is depicted in Fig. 6.6 for a layer representative of the troposphere. At the lower boundary, an individual parcel moves horizontally long enough for heat transfer to occur. Isobaric heat absorption (e.g., through absorption of LW radiation and transfer of sensible heat from the surface) increases the parcel's potential temperature (Fig. 6.7a). The accompanying increase of temperature (Fig. 6.7b) causes the parcel to become positively buoyant and rise. If the time scale of vertical motion is short compared with that of heat transfer, the parcel ascends along an adiabat in state space. Upon reaching the upper boundary, it again moves horizontally. There, isobaric heat rejection (e.g., through LW emission to space) decreases the potential temperature and temperature of the parcel, which therefore becomes negatively buoyant. The parcel then sinks along a different adiabat until it has returned to the surface and completed a thermodynamic cycle.

If it remains unsaturated (e.g., below the LCL), each air parcel comprising the layer preserves its potential temperature away from the boundaries. Consequently, as it traverses the layer, each parcel traces out a uniform profile of θ . The horizontal-mean potential temperature $\bar{\theta}(z)$ is equivalent to an average over all such parcels at

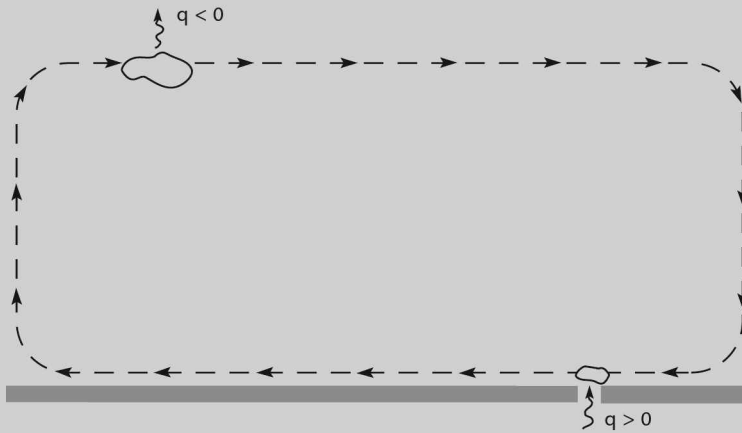


Figure 6.6 Idealized circuit followed by an air parcel during which it absorbs heat at the base of a layer and rejects heat at its top, with adiabatic vertical motion between.

a given elevation. Therefore, it too is independent of height:

$$\bar{\theta}(z) = \text{const.} \quad (6.27.1)$$

As is true for an individual parcel, a uniform vertical distribution of potential temperature corresponds to the constant environmental lapse rate (Sec. 2.4.2)

$$\bar{\Gamma} = \Gamma_d. \quad (6.27.2)$$

Consequently, the thermal structure (6.24) may be regarded as the horizontal mean of a layer in which air is being efficiently overturned.

This interpretation may be extended to saturated conditions (e.g., inside a layer of cloud). An air parcel's evolution is then pseudo-adiabatic, so its equivalent potential temperature is conserved. Efficient overturning of air will make the horizontal-mean equivalent potential temperature $\bar{\theta}_e(z)$ independent of height

$$\bar{\theta}_e(z) = \text{const.} \quad (6.28.1)$$

From Sec. 5.4.3, a uniform vertical distribution of $\bar{\theta}_e(z)$ corresponds to a mean environmental lapse rate

$$\bar{\Gamma} = \Gamma_s. \quad (6.28.2)$$

The corresponding stratification is termed *saturated adiabatic*.

Even though all parcels inside the layer evolve in like fashion, local behavior will differ from the horizontal mean - because conserved properties may still vary from one parcel to another (e.g., due to different histories experienced by those parcels at the layer's boundaries). An ascending parcel will have values

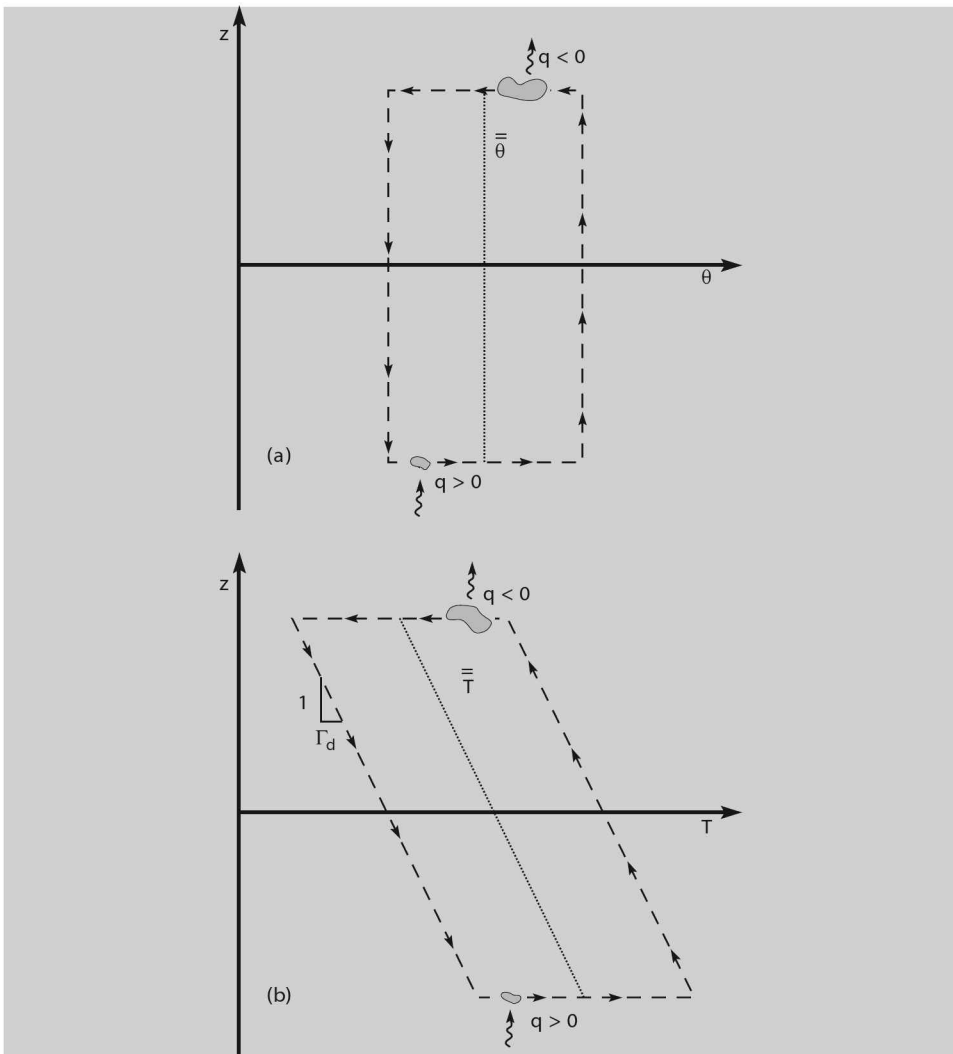


Figure 6.7 Thermodynamic cycle followed by the air parcel in Fig. 6.6 in terms of (a) potential temperature and (b) temperature. Horizontally averaged behavior for a layer comprised of many such parcels also indicated (dotted).

of θ and θ_e that are greater than horizontal-mean values because that parcel will have recently absorbed heat at the lower boundary. Conversely, a descending parcel will have values of θ and θ_e that are less than horizontal-mean values because that parcel will have recently rejected heat at the upper boundary. The actual stratification of the layer at any instant will therefore vary with position.

Were heat transfer at the boundaries eliminated, buoyantly driven motion would “spin down” through turbulent and molecular diffusion. Because diffusion destroys gradients between individual parcels, this limiting state is characterized

by homogeneous distributions of θ and θ_e . It thus corresponds to the environmental lapse rate $\Gamma = \Gamma_d$ or Γ_s . The limiting homogeneous state therefore has the same stratification as the horizontal-mean stratification of the layer that is being efficiently overturned. The latter may therefore be regarded as “statistically well-mixed.”

The stratification (6.28) is characteristic of the troposphere. Close to the saturated adiabatic lapse rate (Fig. 1.2), mean thermal structure in the troposphere reflects efficient overturning of moist air by cumulus and sloping convection. Air ascends at the saturated adiabatic lapse rate Γ_s (~ 5.5 K/km). After being dehydrated through condensation and precipitation, it descends at the dry adiabatic lapse rate Γ_d (~ 10 K/km). The observed global-mean lapse rate in the troposphere, $\bar{\Gamma} \cong 6.5$ K/km, falls between the two, close to Γ_s . Air overturned by convection is continually re-humidified through contact with warm ocean surface. Because convective motion operates on time scales of a day or shorter, it makes the troposphere statistically well-mixed.

In traversing the circuit, the parcel in Fig. 6.7 absorbs heat at high temperature and rejects heat at low temperature. By the Second Law (Sec. 3.2), net heat is absorbed over a cycle. Then the First Law (2.13) implies that the parcel performs net work during its traversal of the circuit. An individual parcel in the above circulation therefore behaves as a heat engine. In fact, the thermodynamic cycle in Fig. 6.7 is analogous to the Carnot cycle pictured in Fig. 3.2, except that heat transfer occurs isobarically instead of isothermally. More expansion work is performed by the parcel during ascent than is performed on the parcel during descent. Net work performed during the cycle is reflected in the area circumscribed by the parcel’s evolution in Fig. 6.7b.

Net heat absorption and work performed by individual air parcels make the general circulation of the troposphere behave as a heat engine, one that is driven thermally by heat transfer at its lower and upper boundaries. Work performed by individual parcels is associated with a redistribution of mass: Air that is effectively warmer and lighter at the lower boundary is exchanged with air that is effectively cooler and heavier at the upper boundary. This redistribution of mass represents a conversion of potential energy into kinetic energy. The conversion of energy maintains the general circulation against frictional dissipation.

6.5.2 Diabatic stratification: A paradigm of the stratosphere

The idealized behavior just described relies on heat transfer being confined to the lower and upper boundaries of the layer, where an air parcel resides long enough for diabatic influences to become important. Between the boundaries, the time scale of motion is short. For motion that operates on longer time scales, typical of the stratosphere, the evolution of an individual air parcel is not adiabatic.

Radiative transfer is the primary diabatic influence outside the boundary layer and cloud. It is characterized by cooling rates of order 1 K day^{-1} in the troposphere (see Fig. 8.24). Cooling rates as large as 10 K day^{-1} occur in the stratosphere and near cloud (Fig. 9.36).

For comparison, the cooling rate associated with adiabatic expansion follows from (2.33) as

$$\begin{aligned}\frac{dT}{dt}\Big|_{ad} &= -\Gamma' \frac{dz}{dt} \\ &= -\Gamma' w,\end{aligned}\tag{6.29.1}$$

where T and z refer to an individual parcel, Γ' denotes its lapse rate, and

$$w = \frac{dz}{dt}\tag{2.29.2}$$

is the parcel's vertical velocity. Cumulus and sloping convection are characterized by a vertical velocity of 0.1 m s^{-1} . Under unsaturated conditions, this gives an adiabatic cooling rate of $\frac{dT}{dt}\Big|_{ad} = 84 \text{ K day}^{-1}$. Even under saturated conditions, the pseudo-adiabatic cooling rate is of order 50 K day^{-1} . Both are much greater than radiative cooling rates. Therefore, heat transfer in cumulus and sloping convection can, to good approximation, be neglected. On the other hand, motion in the stratosphere is characterized by a vertical velocity of 1 mm s^{-1} . This gives an adiabatic cooling rate of only $\frac{dT}{dt}\Big|_{ad} = 0.84 \text{ K day}^{-1}$. For such motion, heat transfer cannot be neglected. Radiative influences may even dominate temperature changes that are introduced by expansion and compression.

Under these circumstances, an air parcel moving vertically interacts thermally with its environment. Because θ is no longer conserved, diabatic motion implies different thermal structure. Consider a layer characterized by

$$\Gamma_d > \Gamma = \text{const.}\tag{6.30.1}$$

Such stratification is termed *subadiabatic*: Temperature decreases with height slower than dry adiabatic. By (6.18), θ increases with height (cf. Fig. 6.5). Then (6.25), together with (2.37), gives

$$\begin{aligned}c &> 0 \\ \delta q > 0 &\quad dT > 0 \\ \delta q < 0 &\quad dT < 0\end{aligned}\tag{6.30.2}$$

for $\Gamma < 0$ (temperature increasing with height). For $\Gamma > 0$ (temperature decreasing with height), c and δq satisfy reversed inequalities. Both imply that an air parcel absorbs heat during ascent, when it experiences expansional cooling, and rejects heat during descent, when it experiences compressional warming. The parcel's potential temperature therefore increases with height. As illustrated in Fig. 6.8, heat is rejected at high temperature at the upper boundary and absorbed at low temperature at the lower boundary. By the Second Law, net heat is rejected over a cycle. Then the First Law implies that net work must be performed on the parcel for it to traverse the circuit.

Opposite to the troposphere, this behavior is characteristic of the stratosphere. Net heat rejection and work performed on individual air parcels make the general circulation of the stratosphere behave as a refrigerator. It is driven mechanically by waves that propagate upward from the troposphere. By transferring momentum, planetary waves exert an influence on the stratosphere analogous to paddle work.

When dissipated, they drive air toward the winter pole, where it experiences radiative cooling (see Figs. 18.11; 8.27). Individual parcels then sink across isentropic surfaces to lower θ (Fig. 18.10). Conversely, air at low latitude experiences radiative warming. Parcels there rise across isentropic surfaces to higher θ .

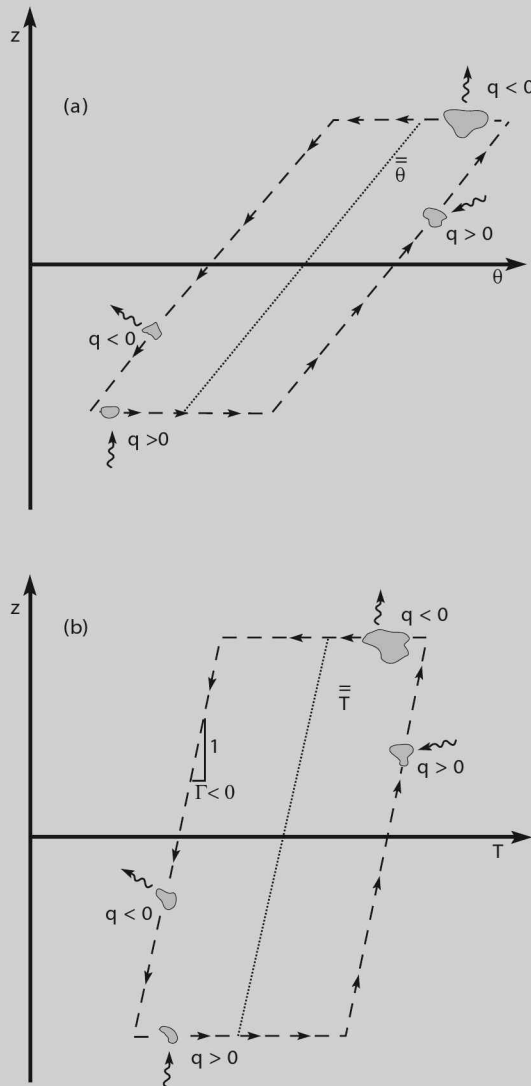


Figure 6.8 As in Fig. 6.7, but for an air parcel whose vertical motion is diabatic and whose temperature increases with height.

For both, vertical motion is slow enough to make diabatic influences important. It enables the temperature of individual parcels to increase with height. The gradual nature of vertical motion in the stratosphere reflects its opposition by

buoyancy. Unlike the troposphere, buoyancy in the stratosphere opposes vertical motion because, invariably, warm (high- θ) air overlies cool (low- θ) air. To exchange effectively-heavier air at lower levels with effectively-lighter air at upper levels, work must be performed against the opposition of buoyancy. The rearrangement of mass represents a conversion of kinetic energy (that of the waves driving the motion) into potential energy. Manifest in temperature, the potential energy is dissipated thermally through LW emission to space.

SUGGESTED REFERENCE

Atmospheric Thermodynamics (1981) by Iribarne and Godson contains a thorough discussion of atmospheric statics.

PROBLEMS

1. Evaluate the discrepancy between geometric and geopotential height, exclusive of anisotropic contributions to gravity, up to 100 km.
2. Show that an atmosphere of uniform negative lapse rate need have no upper bound.
3. For reasons developed in Chap. 12, large-scale horizontal motion is nearly tangential to contours of geopotential height, approximated by the geostrophic velocity

$$\mathbf{v}_g = \frac{1}{f} \left(-\frac{\partial \Phi}{\partial y}, \frac{\partial \Phi}{\partial x} \right),$$

where $\mathbf{v} = (u, v)$ denotes the horizontal component of motion in the eastward (x) and northward (y) directions and Φ is the geopotential along an isobaric surface. Suppose that satellite measurements provide the 3-dimensional distributions of temperature and of the mixing ratio of a long-lived chemical species. From those measurements, describe an algorithm to determine 3-dimensional air motion on large scales at heights great enough to ignore the variation of surface pressure.

4. An upper-level depression, typical of extratropical cyclones during amplification, affects isobaric surfaces aloft but leaves the surface pressure distribution undisturbed. Consider the depression of isobaric height associated with the temperature distribution

$$T(\lambda, \phi, \xi) = \overline{\overline{T}} + \overline{T} \cos(\phi) + T' e^{-\frac{\lambda^2 + (\phi - \phi_0)^2}{L^2}} \cdot f(\xi),$$

where $\overline{\overline{T}}$, \overline{T} , and T' reflect global-mean, zonal-mean, and disturbance contributions to temperature, respectively, $\xi = -\ln(\frac{p}{p_0})$, with $p_0 = 1000$ hPa, is a measure of elevation,

$$f(\xi) = \begin{cases} -\cos\left(\pi \frac{\xi}{\xi_T}\right) & \xi \leq \xi_T \\ 0 & \xi > \xi_T, \end{cases}$$

with $\xi_T = -\ln(\frac{100}{1000})$ reflecting the tropopause elevation, the surface pressure $p_s = p_0$ remains constant, $\phi_0 = \frac{\pi}{4}$, $L = 0.25$ rads, and $\overline{\overline{T}} = 250$ K, $\overline{T} = 30$ K, and $T' = 5$ K.

- (a) Derive an expression for the 3-dimensional distribution of geopotential height.
- (b) Plot the vertical profiles of disturbance temperature and height in the center

of the anomaly. (c) Plot the horizontal distribution of geopotential height for the mid-tropospheric level $\xi = \frac{\xi_T}{2}$, as a function of λ and $\phi > 0$. (d) Plot a vertical section at latitude ϕ_0 , showing the elevations ξ of isobaric surfaces as functions of λ .

5. Under the conditions of (4), how large must the anomalous temperature T' be for the disturbance to form a cut-off low at the mid-tropospheric level $\xi = \frac{\xi_T}{2}$?
6. Unlike mid-latitude depressions, hurricanes are warm-core cyclones that are invariably accompanied by a low in surface pressure. Consider a hurricane which is associated with the temperature distribution

$$T(\lambda, \phi, \xi) = \bar{T} + T'(\lambda, \phi)e^{-\alpha\xi},$$

where \bar{T} and T' reflect zonal-mean and disturbance contributions to temperature, $\xi = -\ln\left(\frac{p}{p_0}\right)$, with $p_0 = 1000$ hPa, is a measure of elevation,

$$T'(\lambda, \phi) = \bar{T}\alpha\xi_s(\lambda, \phi)e^{\alpha\xi_s(\lambda, \phi)},$$

in which

$$\xi_s(\lambda, \phi) = 0.1e^{-\frac{\lambda^2 + \phi^2}{L^2}},$$

with $\alpha^{-1} = \frac{\xi_T}{2}$, and $\xi_T = \ln\left(\frac{100}{1000}\right)$ reflect the surface and tropopause pressures, respectively, and $L = 0.1$ rads. (a) Derive an expression for the 3-dimensional distribution of geopotential height. (b) Plot the vertical profiles of disturbance temperature and height in the center of the anomaly. (c) Plot the horizontal distribution of geopotential height for the mid-tropospheric level $\xi = \frac{\xi_T}{2}$. (d) Plot a vertical section at the equator, showing the elevations ξ of isobaric surfaces as functions of λ .

7. A surface analysis plots pressure adjusted to a common reference level: mean sea level (MSL). Different locations can then be compared meaningfully. Derive an approximate expression for the pressure at MSL in terms of the local surface height z_s , surface pressure p_s , and the global-mean surface temperature $\bar{T} = 288$ K. (b) A surface analysis shows a ridge with a maximum of 1040 hPa over Denver, which lies at an elevation of 5300' above MSL. What surface pressure was actually recorded there?
8. Deep convection inside a region of the tropics releases latent heat to achieve a column heating rate of 250 W m^{-2} . (a) If, uncompensated, that heating would produce a temperature perturbation that is invariant with height up to the tropopause, calculate the implied change of thickness for the 1000 hPa–100 hPa layer after 1 day (cf. Fig. 7.8). (b) What process, in reality, compensates the tendency of tropospheric thickness implied in (a)?
9. Show that the polytropic potential temperature in Prob. 2.16 is conserved under the diabatic conditions described in Sec. 6.5.2.
10. Use (2.27), with z replaced by $-\ln p$, to relate the area circumscribed in Fig. 6.7 to the work performed by the air parcel during one thermodynamic cycle.

11. Approximate the thermodynamic cycle in Fig. 6.7 by one comprised of two isobaric legs and two adiabatic legs to show that (a) net work is performed during a complete cycle and (b) the work performed is proportional to the change of temperature associated with isobaric heat transfer.
12. Show that sub-adiabatic stratification requires an individual parcel moving vertically to, in general, absorb (reject) heat as it ascends (descends).

Static stability

The property responsible for the distinctly different character of the troposphere and stratosphere is vertical stability. Because it applies even under motionless conditions and because displacements of air remain in hydrostatic balance, vertical stability is also referred to as *static stability*. Although forces are never far from hydrostatic equilibrium, vertical motions are introduced by forced lifting over elevated terrain and through buoyancy. Buoyantly driven motion is related closely to the stability of the atmospheric mass distribution. The latter, in turn, is shaped by transfers of energy between the Earth's surface, the atmosphere, and deep space. By promoting convection in some regions and suppressing it in others, vertical stability controls a wide range of properties.

7.1 REACTION TO VERTICAL DISPLACEMENT

Because of hydrostatic equilibrium, together with the compressibility of air, density decreases upward regardless of temperature structure (Fig. 6.5b). Thus, mean stratification invariably has lighter air configured over heavier air. This suggests stability with respect to vertical displacements. Were air incompressible, this would indeed be the case.

Consider an arbitrary air parcel, inside the layer pictured in Fig. 7.1. In a linear stability analysis, this parcel will be used to establish how the layer reacts to infinitesimal disturbances from equilibrium. Although the analysis focuses on an individual parcel, stability actually refers to the layer as a whole.

Suppose the parcel is disturbed by a virtual displacement $\delta z' = z'$, where primes distinguish properties of the parcel from those of its environment. Through expansion or compression, the parcel automatically adjusts to the environmental pressure to

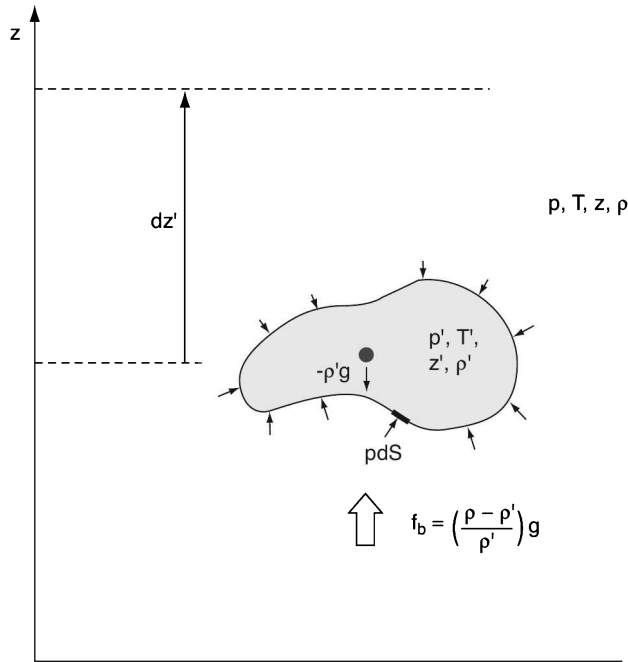


Figure 7.1 Schematic of an air parcel, the properties of which are distinguished by primes, inside a designated layer. The specific buoyancy force f_b results from an imbalance between the parcel's weight and the net pressure force acting over its surface.

preserve mechanical equilibrium. Thus

$$p' = p$$

throughout the parcel's displacement. Further, so long as the displacement occurs on a time scale short compared with that of heat transfer with the environment, the parcel's evolution may be regarded as adiabatic.

Per unit volume, Newton's second law for the parcel is

$$\rho' \frac{d^2 z'}{dt^2} = -\rho' g - \frac{\partial p'}{\partial z'}, \quad (7.1.1)$$

which follows from a development analogous to the one leading to (1.16). For the environment, the momentum balance is simply hydrostatic equilibrium:

$$0 = \rho g - \frac{\partial p}{\partial z}. \quad (7.1.2)$$

Subtracting (7.1) and incorporating mechanical equilibrium yields for the parcel

$$\rho' \frac{d^2 z'}{dt^2} = (\rho - \rho') g. \quad (7.2.1)$$

This relationship also follows from Archimedes' principle (Prob. 7.3). Per unit mass, the parcel's momentum balance is then described by

$$\begin{aligned}\frac{d^2 z'}{dt^2} &= \left(\frac{\rho - \rho'}{\rho'} \right) g \\ &= f_b,\end{aligned}\tag{7.2.2}$$

where f_b is the (specific) net buoyancy force experienced by the parcel. With (7.1), the buoyancy force is recognized as an imbalance between the parcel's weight and the vertical component of pressure force acting over its surface.

Consider how the displaced parcel's temperature varies in relation to that of its environment. With the gas law, (7.2.2) may be expressed

$$\frac{d^2 z'}{dt^2} = g \left(\frac{T' - T}{T} \right).\tag{7.3}$$

For small displacements from its undisturbed elevation, the parcel's temperature decreases with height at the parcel lapse rate: $\Gamma' = \Gamma_d$ (Γ_s) under unsaturated (saturated) conditions. Thus

$$T' = T_0 - \Gamma' z'\tag{7.4.1}$$

describes the first-order variation of the parcel's temperature from its undisturbed value T_0 , which equals the environmental temperature at the parcel's undisturbed height. On the other hand, the temperature of the environment decreases with height at the local environmental lapse rate Γ , so the corresponding variation of environmental temperature is

$$T = T_0 - \Gamma z'.\tag{7.4.2}$$

The parcel's lapse rate is determined by its thermodynamics. Under adiabatic conditions, Γ' is independent of the environmental lapse rate Γ - because the parcel is then thermally isolated. The environmental lapse rate, on the other hand, is determined by the history of air residing in the layer (e.g., by where that air has been and what thermodynamic influences it has experienced).

Incorporating (7.4) transforms the vertical momentum balance (7.3) into

$$\begin{aligned}\frac{d^2 z'}{dt^2} &= \frac{g}{T} (\Gamma - \Gamma') z' \\ &= f_b.\end{aligned}\tag{7.5}$$

Hence, the specific buoyancy force experienced by the parcel is proportional to its displacement z' and to the difference between lapse rates of the parcel and the environment.

7.2 STABILITY CATEGORIES

If the parcel is unsaturated,

$$\Gamma' = \Gamma_d.\tag{7.6.1}$$

Because Γ' is then constant, (7.4.1) is also valid for finite displacements. If the parcel is saturated,

$$\Gamma' = \Gamma_s,\tag{7.6.2}$$

where Γ_s implies its value at the undisturbed height.

STABILITY CATEGORIES

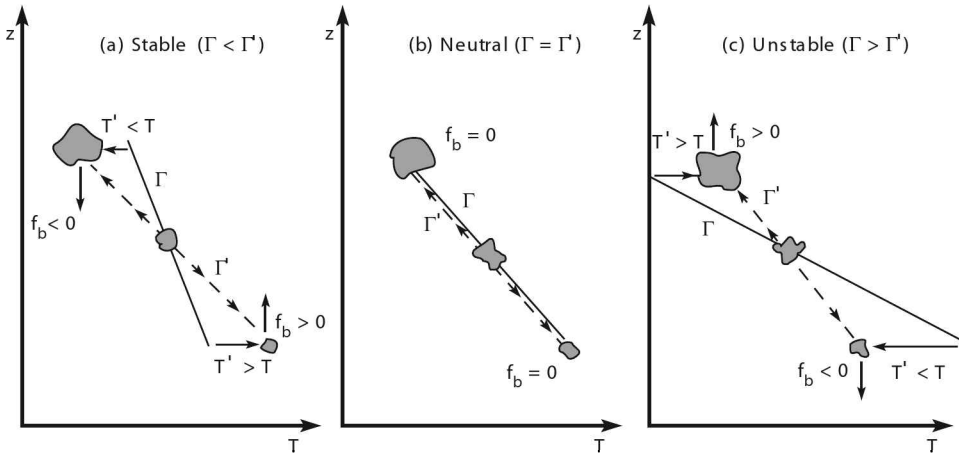


Figure 7.2 Buoyancy reaction experienced by a displaced air parcel, in terms of the environmental lapse rate Γ and the parcel lapse rate Γ' , which equals Γ_d (Γ_s) under unsaturated (saturated) conditions. (a) $\Gamma < \Gamma'$, (b) $\Gamma = \Gamma'$, (c) $\Gamma > \Gamma'$.

7.2.1 Stability in terms of temperature

For either of the above conditions, three possibilities exist (Fig. 7.2):

- (1) $\Gamma < \Gamma'$: Environmental temperature decreases with height slower than the displaced parcel's temperature. Γ is *subadiabatic*. Then (7.5) implies

$$\frac{1}{z'} \frac{d^2 z'}{dt^2} = \frac{f_b}{z'} < 0.$$

The parcel experiences a buoyancy force f_b that opposes its displacement z' . If it is displaced upward, the parcel becomes heavier than its environment and thus negatively buoyant. Conversely, if it is displaced downward, the parcel becomes lighter than its environment and positively buoyant. This buoyancy reaction constitutes a *positive restoring force*, one that acts to restore the system to its undisturbed height - irrespective of the sense of the displacement. The atmospheric layer is then said to be *statically stable*. The environmental temperature profile is said to possess *positive stability*.

- (2) $\Gamma = \Gamma'$: Environmental temperature decreases with height at the same rate as the parcel's temperature. Γ is *adiabatic*. Under these circumstances,

$$\frac{1}{z'} \frac{d^2 z'}{dt^2} = \frac{f_b}{z'} = 0,$$

so the buoyancy force f_b vanishes and the displaced parcel experiences no restoring force. The atmospheric layer is then said to be *statically neutral*. The environmental temperature profile is said to possess *zero stability*.

- (3) $\Gamma > \Gamma'$: Environmental temperature decreases with height faster than the parcel's temperature. Γ is *superadiabatic*. Then

$$\frac{1}{z'} \frac{d^2 z'}{dt^2} = \frac{f_b}{z'} > 0.$$

The parcel experiences a buoyancy force f_b that reinforces its displacement z' . If it is displaced upward, the parcel becomes lighter than its environment and thus positively buoyant. Conversely, if it is displaced downward, the parcel becomes heavier than its environment and negatively buoyant. This buoyancy reaction constitutes a *negative restoring force*, one that drives the system away from its undisturbed height – irrespective of the sense of the displacement. The atmospheric layer is then said to be *statically unstable*. The temperature profile is said to possess *negative stability*.

Even though density decreases upward in all three cases, through buoyancy, each leads to a very different response. The layer's stability is not determined by its vertical profile of density because the density of a displaced air parcel changes through expansion and compression. Depending on whether its density decreases with height faster or slower than environmental density, a displaced parcel will find itself heavier or lighter than its environment. It will, therefore, experience a restoring force that is either positive or negative.

The degree of stability or instability is reflected in the magnitude of the restoring force. The latter is proportional to the difference between the lapse rates of the environment and the parcel. According to (7.5), the parcel lapse rate Γ' serves as a reference against which the local environmental lapse rate Γ is measured to determine stability (Fig. 7.3). Under unsaturated conditions, $\Gamma' = \Gamma_d$. If environmental temperature decreases with height slower than Γ_d (i.e., Γ is subadiabatic), the layer is stable. If environmental temperature decreases with height at a rate equal to Γ_d (i.e., Γ is adiabatic), the layer is neutral. If environmental temperature decreases with height faster than Γ_d (i.e., Γ is superadiabatic), the layer is unstable.

Under saturated conditions (e.g., inside a layer of stratiform cloud), the same criteria hold, but with Γ_s in place of Γ_d . Because

$$\Gamma_s < \Gamma_d,$$

Fig. 7.3 implies that static stability is violated more easily under saturated conditions than under unsaturated conditions. That is, the range of stable lapse rate is narrower under saturated conditions than under unsaturated conditions. Reduced stability of saturated air follows from the release (absorption) of latent heat, which warms (cools) an ascending (descending) parcel. By adding positive (negative) buoyancy, the release (absorption) of latent heat reinforces vertical displacements, destabilizing the layer.

7.2.2 Stability in terms of potential temperature

Stability criteria can also be expressed in terms of potential temperature. For unsaturated conditions, logarithmic differentiation of (2.31) gives

$$\frac{1}{\theta} \frac{d\theta}{dz} = \frac{1}{T} \frac{dT}{dz} - \frac{\kappa}{p} \frac{dp}{dz}.$$

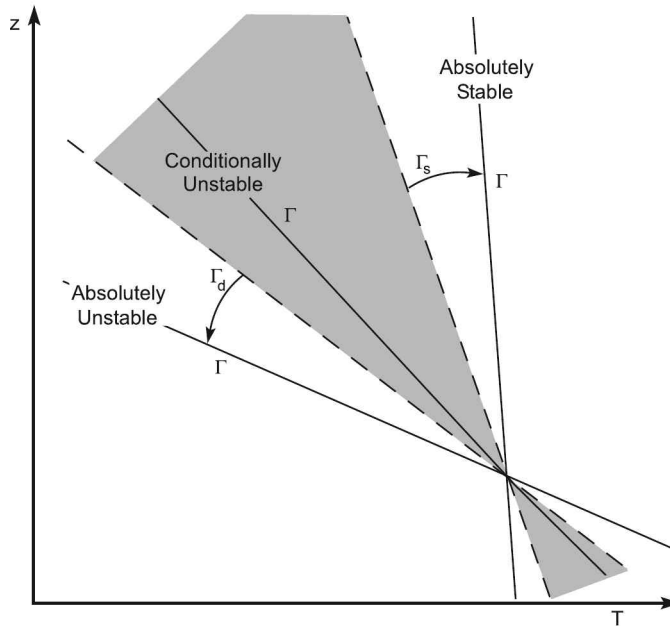


Figure 7.3 Vertical stability in terms of temperature, determined by the environmental lapse rate Γ in relation to the adiabatic lapse rate Γ' .

Hydrostatic equilibrium and the gas law transform this into

$$\frac{1}{\theta} \frac{d\theta}{dz} = \frac{1}{T} \frac{dT}{dz} + \frac{\kappa}{H}.$$

Then with (1.17), the vertical gradient of potential temperature may be expressed in terms of the lapse rate

$$\frac{d\theta}{dz} = \frac{\theta}{T} (\Gamma_d - \Gamma). \quad (7.7)$$

As it follows from the definitions of potential temperature and lapse rate, (7.7) also governs the environment. An environmental lapse rate $\Gamma < \Gamma_d$ (subadiabatic) corresponds to $\frac{d\theta}{dz} > 0$. Potential temperature then increases upward. An environmental lapse rate $\Gamma = \Gamma_d$ (adiabatic) corresponds to $\frac{d\theta}{dz} = 0$. The profile of potential temperature is then uniform. An environmental lapse rate $\Gamma > \Gamma_d$ (superadiabatic) corresponds to $\frac{d\theta}{dz} < 0$. Potential temperature then decreases upward.

Thus, for unsaturated conditions, the stability criteria established previously in terms of temperature translate into the simpler criteria in terms of potential temperature:

$$\begin{aligned} \frac{d\theta}{dz} > 0 & \quad (\text{stable}) \\ \frac{d\theta}{dz} = 0 & \quad (\text{neutral}) \\ \frac{d\theta}{dz} < 0 & \quad (\text{unstable}). \end{aligned} \quad (7.8)$$

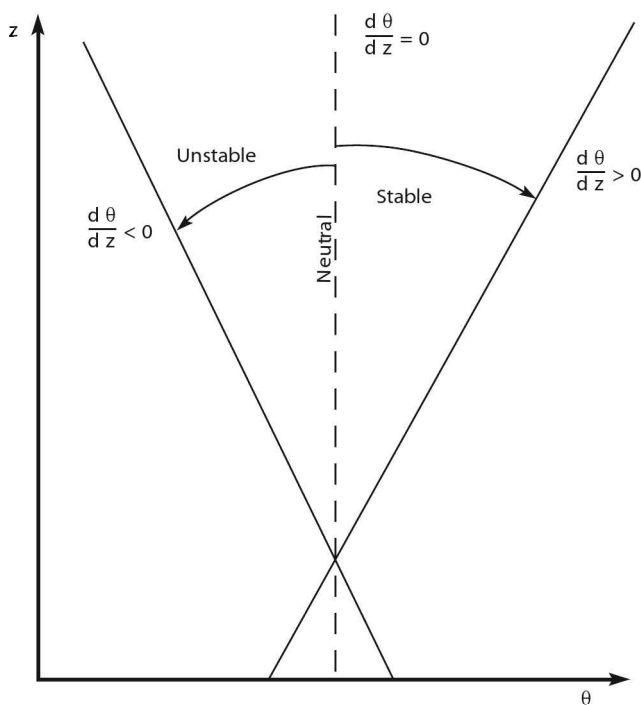


Figure 7.4 Vertical stability, under unsaturated conditions, in terms of potential temperature.

Illustrated in Fig. 7.4, these criteria have the same form as those expressed in terms of temperature. However, the reference profile corresponding to neutral stability is now just a uniform distribution of θ . This has a parallel in the vertical stability of water, which is incompressible. If potential temperature increases with height, air that is effectively warmer and lighter (with account of compressibility) overlies air that is effectively colder and heavier. The layer is then stable. If potential temperature is uniform with height, the layer is neutral. If θ decreases with height, air that is effectively colder and heavier overlies air that is effectively warmer and lighter. The layer is then unstable.

The simplified stability criteria (7.8) reflect the close relationship between θ and atmospheric motion. Recall that potential temperature is conserved for individual air parcels because it accounts for expansion work. Therefore, a displaced parcel traces out a uniform profile of θ ; it defines the reference thermal structure corresponding to neutral stability. The same is true of temperature for an incompressible fluid. It is for this reason that potential temperature plays a role for air analogous to the role that temperature plays for water. Hence, the vertical distribution of θ describes the effective stratification of mass when compressibility is taken into account.

Under saturated conditions, θ is not conserved. However, θ_e is conserved. It bears a relationship to Γ_s analogous to the role that θ bears to Γ_d under unsaturated conditions. Using arguments similar to that above, it may be shown that the criteria for stability under saturated conditions are also given by (7.8), but with θ_e in place of θ

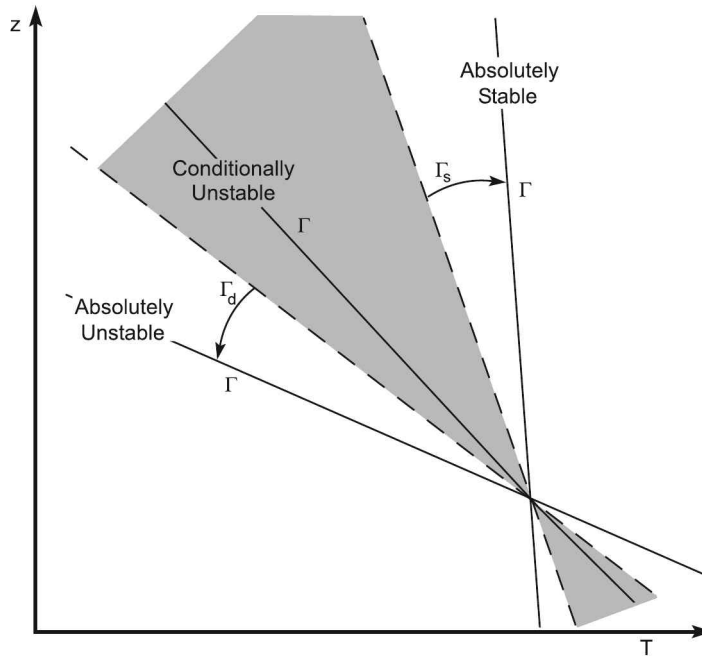


Figure 7.5 Vertical stability of moist air in terms of temperature and the environmental lapse rate Γ .

(Prob. 7.16). Thus, the vertical distribution of θ_e describes the effective stratification of mass, when both compressibility and latent heat release are taken into account.

7.2.3 Moisture dependence

Humidity complicates the description of vertical stability by introducing the need for two categories: One is valid under unsaturated conditions. The other is valid under saturated conditions, wherein the release of latent heat must be accounted for. Because $\Gamma_s < \Gamma_d$, three possibilities exist (Fig. 7.5). If

$$\Gamma < \Gamma_s, \quad (7.9.1)$$

a layer is *absolutely stable*. A small displacement of air will be met by a positive restoring force - irrespective of whether the air is saturated, because (7.9.1) ensures that a displaced parcel will cool (warm) faster than its environment. If

$$\Gamma > \Gamma_d, \quad (7.9.2)$$

the layer is *absolutely unstable*. A small displacement of air will result in a negative restoring force - irrespective of whether the air is saturated, because (7.9.2) ensures that a displaced parcel will cool (warm) slower than its environment. If

$$\Gamma_s < \Gamma < \Gamma_d, \quad (7.9.3)$$

the layer is *conditionally unstable*. A small displacement of air will then result in a positive restoring force if the layer is unsaturated, but a negative restoring force if it is saturated.

The first two possibilities apply irrespective of whether or not the parcel is saturated – because the restoring force does not change sign. The third possibility, however, hinges on the parcel's state. For finite displacements, an air parcel that is initially unsaturated can become saturated and vice versa. If the layer is conditionally unstable, the restoring force experienced by such a parcel will change sign. Such conditions are treated in Sec. 7.4.

7.3 IMPLICATIONS FOR VERTICAL MOTION

The stability of a layer determines its ability to support vertical motion and thus to support transfers of heat, momentum, and constituents. To conserve mass, vertical motion must be compensated by horizontal motion (cf. Fig. 5.4). Hydrostatic stability therefore also influences horizontal transport. Three-dimensional (3D) turbulence, which disperses atmospheric constituents, involves both vertical and horizontal motion. Consequently, suppressing vertical motion simultaneously suppresses the horizontal component of 3D eddy motion and hence turbulent dispersion.

A layer that is stably stratified inhibits vertical motion. Small vertical displacements, introduced mechanically by flow over elevated terrain or thermally through isolated heating, are then opposed by the positive restoring force of buoyancy (Prob. 7.5). Conversely, a layer that is unstably stratified promotes vertical motion through the negative restoring force of buoyancy. Work performed against buoyancy under stable conditions or by buoyancy under unstable conditions reflects a conversion between potential energy and kinetic energy. It controls the layer's evolution.

The vertical momentum balance for an unsaturated air parcel may be expressed

$$\frac{d^2 z'}{dt^2} + N^2 z' = 0, \quad (7.10.1)$$

where

$$\begin{aligned} N^2 &= \frac{g}{T} (\Gamma_d - \Gamma) \\ &= -\frac{f_b}{z'} \end{aligned} \quad (7.10.2)$$

defines the *buoyancy* or *Brunt-Väisällä frequency* N . Equation (7.10.1) describes a simple harmonic oscillator, with N^2 reflecting the “stiffness” of the buoyancy spring. From (7.7), the Brunt-Väisällä frequency may be expressed simply

$$N^2 = g \frac{d \ln \theta}{dz}. \quad (7.10.3)$$

Proportional to the restoring force of buoyancy, N^2 is a measure of stability.

If a layer is stable, $N^2 > 0$. Then (7.10.1) has solutions of the form

$$z'(t) = A e^{iNt} + B e^{-iNt}. \quad (7.11.1)$$

A displaced parcel oscillates about its undisturbed height, with kinetic energy repeatedly exchanged with potential energy. The stronger the stability (i.e., the stiffer the buoyancy spring), the more rapid is the parcel's oscillation, and the smaller is its

displacement. Owing to the positive restoring force of buoyancy, a small imposed displacement remains small. By limiting vertical motion, positive stability also imposes a constraint on horizontal motion (e.g., suppressing eddy motion associated with 3D turbulence).

Under neutral conditions, $N^2 = 0$ and the restoring force vanishes. Imposed displacements are then met with no opposition. In that event, the solution of (7.10.1) grows linearly with time. It describes a displaced parcel moving inertially, with no conversion between potential and kinetic energy. Under these circumstances, a small imposed displacement ultimately evolves into finite displacements of air. That long-term behavior violates the linear analysis leading to (7.10.1), which is predicated on infinitesimal displacements. However, in the presence of friction, displacements remain bounded (Prob. 7.17).

Under positive and neutral stability, air displacements can remain small enough for the mean stratification of a layer to be preserved. A layer of negative stability evolves very differently. In that event, $N^2 = -\tilde{N}^2 < 0$ and solutions of (7.10.1) have the form

$$z'(t) = Ae^{\tilde{N}t} + Be^{-\tilde{N}t}. \quad (7.11.2)$$

A parcel's displacement then grows exponentially with time through reinforcement by the negative restoring force of buoyancy. The first term in (7.11.2) dominates the long-term behavior, violating the linear analysis that led to (7.10.1). Unlike behavior under positive and neutral stability, air displacements need not remain bounded. Except for small \tilde{N} , displacements amplify exponentially – even in the presence of friction (Prob. 7.18). A small initial disturbance then evolves into fully developed convection. Nonlinear effects eventually limit amplification by modifying the stratification of the layer. By rearranging mass, convective cells alter N^2 and hence the buoyancy force experienced by individual air parcels. The simple linear description (7.10), which relies on mean properties of the layer remaining constant, then breaks down.

Amplifying motion described by (7.11.2) is fueled by a conversion of potential energy, which is associated with the vertical distribution of mass, into kinetic energy, which is associated with convective motion. By feeding off of the layer's potential energy, air motion modifies its stratification. Fully developed convection results in efficient vertical mixing. It rearranges the conserved property θ (θ_e) into a distribution that is statistically homogeneous (Sec. 6.5). According to (7.8), this limiting distribution corresponds to a state of neutral stability. Thus, small disturbances to an unstable layer amplify, eventually evolving into fully developed convection. Convective overturning neutralizes the instability by mixing θ (θ_e) into a uniform distribution. In that limiting state, no more potential energy is available for conversion to kinetic energy. Without a regeneration of instability, convective motion then decays through frictional dissipation.

7.4 FINITE DISPLACEMENTS

Stability criteria established in Sec. 7.2 hold for infinitesimal displacements. A parcel that is initially unsaturated then remains so and likewise for one that is initially saturated. For finite displacements, a disturbed air parcel can evolve between saturated and unsaturated conditions. Such displacements are particularly relevant to conditional instability (7.9.3). A displaced parcel may then experience a restoring force of

one sign initially, but of opposite sign thereafter. If the layer itself is displaced vertically (e.g., in sloping convection or through forced lifting over elevated terrain), finite displacement enables the layer's stability itself to change.

7.4.1 Conditional instability

Consider an unsaturated parcel that is displaced upward inside the conditionally unstable layer shown in Fig. 7.6a. Depicted there is a simple representation of the tropical troposphere, which is assigned a constant and conditionally unstable lapse rate, and the lower stratosphere, which is assigned a lapse rate of zero. Below its LCL, the displaced parcel cools at the dry adiabatic lapse rate Γ_d . It therefore cools more rapidly than its environment (7.9.3). The parcel thus finds itself cooler than its environment and hence negatively buoyant. It experiences a positive restoring force, one that intensifies with height (7.5). Work must therefore be performed against buoyancy to complete the displacement.

If enough work is performed, the parcel will eventually cross its LCL. It then cools slower, due to the release of latent heat, at the saturated adiabatic lapse rate Γ_s . Because the layer is conditionally unstable, the parcel also cools slower than its environment (7.9.3). Consequently, the temperature difference between the parcel and its environment narrows. The positive restoring force of buoyancy then weakens with height.

After sufficient displacement, the temperature profile of the parcel crosses the profile of environmental temperature. There, the parcel finds itself warmer than its environment. Positively buoyant, it thereafter experiences a negative restoring force. Buoyancy then performs work on the parcel, which accelerates upward autonomously. The height where this occurs is the *Level of Free Convection* (LFC).

Above the LFC, the parcel accelerates upward until, eventually, its temperature crosses the profile of environmental temperature a second time. The height where this occurs is the Level of Neutral Buoyancy (LNB). Above the LNB, the parcel again finds itself cooler than its environment and thus negatively buoyant. Buoyancy then opposes further ascent. Air that accelerated upward above the LFC therefore diverges horizontally in a neighborhood of the LNB, forming extensive anvil.

Energetics of the displaced parcel provide some insight into deep convective motion. Under adiabatic conditions, the buoyancy force is conservative (e.g., the work performed along a cyclic path vanishes). Therefore, the potential energy P of the displaced parcel may be introduced in a manner similar to that used in Sec. 6.2 to introduce the geopotential

$$\begin{aligned} dP &= \delta w_b \\ &= -f_b dz', \end{aligned} \tag{7.12}$$

where δw_b is the incremental work performed against buoyancy. Defining a reference value of zero potential energy at the undisturbed height z_0 and incorporating (7.2) obtains

$$\begin{aligned} P &= \int_{z_0}^z \left(\frac{v - v'}{v} \right) g dz' \\ &= \int_{p_0}^p (v' - v) dp. \end{aligned} \tag{7.13}$$

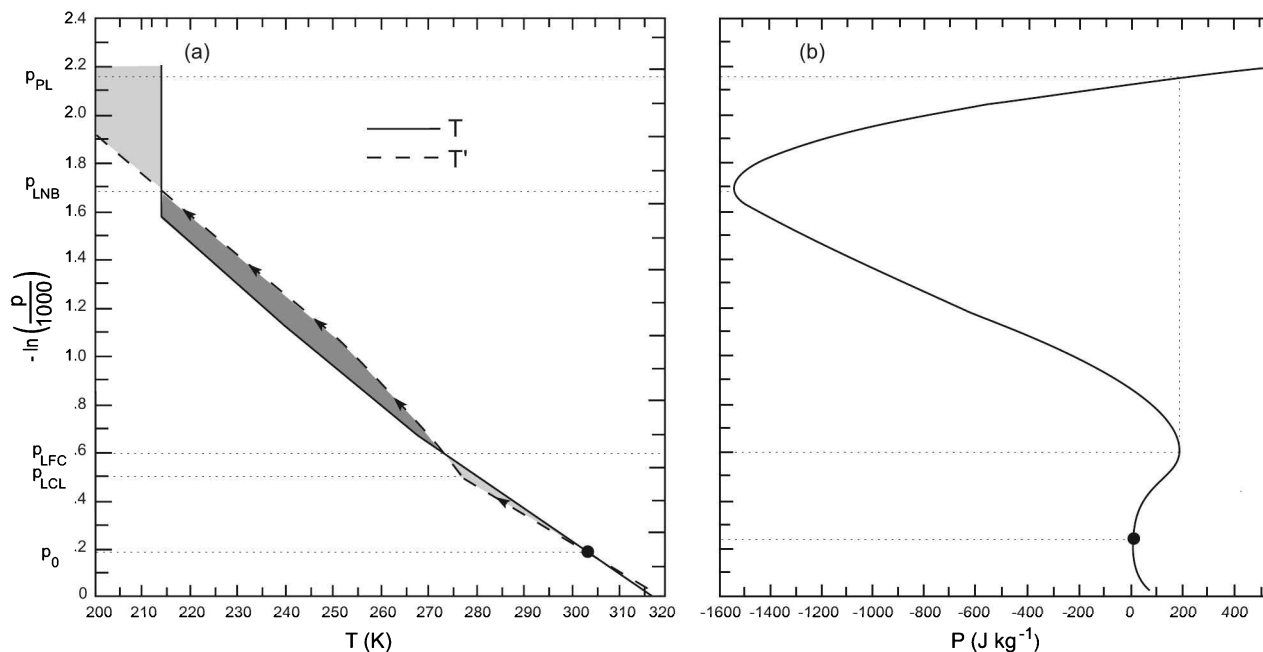


Figure 7.6 (a) Environmental temperature (solid), under conditionally unstable conditions symbolic of the tropical troposphere ($\Gamma \cong 8.5\ K\ km^{-1}$) and stratosphere, and parcel temperature (dashed) displaced from the undisturbed level $p_0 = 800\ hPa$. Parcel temperature decreases with height along a dry adiabat below the LCL and along a saturated adiabat above the LCL. It crosses the profile of environmental temperature at the *level of free convection* p_{LFC} and again at the Level of Neutral Buoyancy p_{LNB} above the tropopause. Not far overhead is the limiting height of convective overshoot, the Penetration Level p_{PL} . (b) Potential energy of the displaced parcel, which is proportional to minus the cumulative area $A(p)$ under the parcel's temperature profile and above the environmental temperature profile that is shaded in (a).

Then, with the gas law, the parcel's potential energy becomes

$$\begin{aligned} P(p) &= -R \int_p^{p_0} (T' - T) d \ln p \\ &= -RA(p), \end{aligned} \quad (7.14)$$

where $A(p)$ is the cumulative area between the temperature profile of the parcel T' and that of the environment T to the level p (Fig. 7.6a).

Above the undisturbed level but below the LFC, $T' < T$. P then increases upward (Fig. 7.6b). Beneath the undisturbed level, T' and T are reversed. Hence P increases downward as well. Increasing P away from the undisturbed level represents a "potential well" in which the parcel is bound. Air beneath the LFC is thus trapped. It can be released from the potential well only if work is performed against the positive restoring force of buoyancy. Achieving that condition requires finite work and hence finite vertical displacement.

If sufficient work is performed to lift the parcel to its LFC, then P reaches a maximum. Thereafter, $T' > T$, so P decreases upward. Under conservative conditions,

$$\Delta K = -\Delta P, \quad (7.15.1)$$

which follows from (7.2). Decreasing P above the LFC represents a conversion of potential energy to kinetic energy. Buoyancy work accelerates the parcel upward, driving deep convection.

The total potential energy available for conversion to kinetic energy is termed the *Convective Available Potential Energy* (CAPE). Represented by the dark-shaded area in Fig. 7.6a, it equals the total work performed by buoyancy above the LFC. That, in turn, equals the large drop of potential energy in Fig. 7.6b. Because the parcel's temperature eventually crosses the environmental temperature profile a second time (at the LNB), CAPE is finite. So is the kinetic energy that can be acquired by the parcel. An upper bound on the parcel's kinetic energy follows from (7.15.1) as

$$\begin{aligned} \frac{w'^2}{2} &= P(p_{LFC}) - P(p_{LNB}) \\ &= R \int_{p_{LNB}}^{p_{LFC}} (T' - T) d \ln p \\ &= \text{CAPE}. \end{aligned} \quad (7.15.2)$$

It describes adiabatic ascent under conservative conditions. In practice, however, mixing with the environment makes a cumulus updraft inherently nonconservative (cf. Fig. 5.4). The upward velocity in (7.15.2) is therefore realized only inside the core of a broad updraft, where air is isolated from entrainment and mixing that damp upward motion at the periphery of the updraft.

Above the LNB, T' and T are again reversed. P therefore increases upward above p_{LNB} . Negatively buoyant, the parcel is then bound in another potential well. Despite opposition by buoyancy, the parcel overshoots its new equilibrium level - due to the kinetic energy that it acquired above the LFC. However, P increases sharply above p_{LNB} , because environmental temperature above the tropopause diverges rapidly from the parcel's temperature. Consequently, penetration into the stable layer overhead is shallow compared with the depth traversed through the conditionally unstable layer below. Eventually, the parcel reaches a height where all of the kinetic energy that it

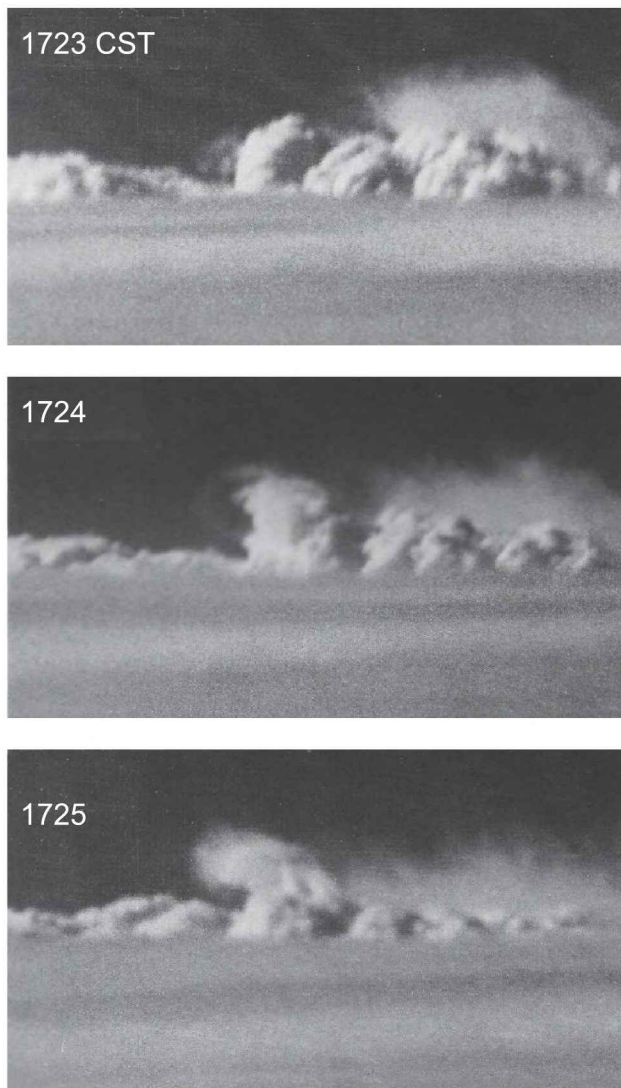


Figure 7.7 Convective domes overshooting the anvil of a cumulonimbus complex, at 2-minute intervals. Fibrous features visible in the second plate mark ice particles that nucleate spontaneously when stratospheric air is displaced by convection. After Fujita (1992).

acquired above the LFC has been re-converted into potential energy. This defines the *Penetration Level* (PL). As the motion is presumed to be conservative, the PL coincides with the height where P equals its previous maximum at the LFC.

The penetration level provides an estimate of the height of “convective overshoots”, which accompany cumulus towers. However, like the maximum updraft in (7.15.2), this estimate is only an upper bound. In practice, some of the kinetic energy driving convective overshoots is siphoned off by mixing with the environment. Figure 7.7

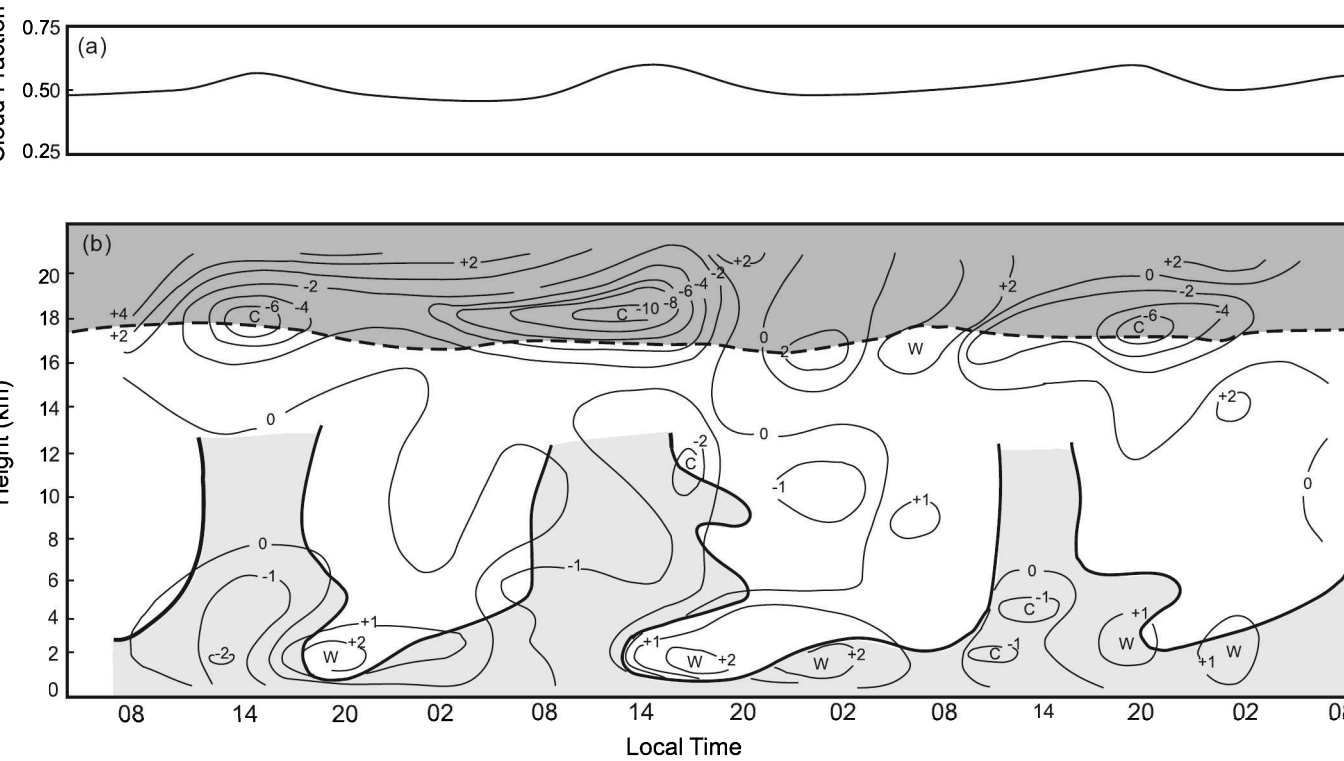


Figure 7.8 Time series of (a) high cloud cover and (b) moisture and temperature derived from radiosondes and satellite IR radiances during the winter MONEX observing campaign. Dark shading indicates the fractional coverage by cold convective cloud anvil. Relative humidity exceeding 80% is shaded light. Temperature departure from the local time-mean value is contoured. Adapted from Johnson and Kriete (1982).

shows an overshooting cumulonimbus complex at 2-minute intervals. Several convective domes have penetrated above the anvil, where air eventually diverges horizontally. Negatively buoyant, overshooting cloud splashes back into the anvil, but not before it has been sheared and mixed with stratospheric air. Thus, convective overshoots lead to an irreversible rearrangement and eventually mixing of tropospheric and stratospheric air.

This feature makes penetrative convection instrumental in processes that control the mean elevation of the tropopause. The tropopause is determined by a competition between vertical overturning, which prevails in the troposphere and drives stratification toward neutral stability, and radiative transfer, which prevails in the stratosphere and drives stratification toward positive stability. Figure 7.8 displays anomalous temperature, moisture, and cloud cover for an area in the tropics, each as a function of time. Towering moisture (light shading in Fig. 7.8b) and the amplification of cold high cloud (Fig. 7.8a) mark episodes of deep cumulus convection. They are contemporaneous with the tropopause (dashed) ascending to higher elevation and, simultaneously, with it becoming anomalously cold (contours) – by as much as 10 K. Jointly, these anomalies reflect an upward extension of positive lapse rate that characterizes the troposphere. It invades levels that previously were stable and in the stratosphere.

7.4.2 Entrainment

In practice, the vertical velocity and penetration level of a cumulus updraft are reduced from adiabatic values by nonconservative effects. Environmental air, which is cooler and drier (low θ_e), is entrained into and mixed with updraft air, which is anomalously warm and moist (high θ_e); cf. Fig. 5.4. Simultaneously, high- θ_e air is detrained from the updraft, mixed with low- θ_e air in the environment.¹ Such mixing depletes ascending parcels of positive buoyancy and kinetic energy. Only in the core of a broad updraft are adiabatic values ever realized. Mixing also modifies the environment of a cumulus updraft. This mediates gradients of temperature and moisture that control buoyancy.

For reasons established in Chap. 9, entrainment of environmental air is intrinsic to cumulus convection. It renders the evolution of ascending parcels diabatic. The impact this process exerts on convection is illustrated by regarding entrained environmental air to become uniformly mixed with ascending air inside a moist updraft. Following Holton (2004), we consider a parcel of mass m inside the updraft pictured in Fig. 5.4. Through mixing, additional mass is incorporated into that parcel. The parcel may then be identified with a certain percentage of the overall mass crossing a given level. If μ is a conserved property (e.g., $\mu = r$ beneath the LCL), then $m\mu'$ represents how much of that property is associated with the parcel under consideration. As before, primes distinguish properties of the parcel from those of the environment.

¹ Detrainment, which refers to the transfer of air to the environment, compensates entrainment that is introduced through mixing (e.g., along the wall of a cumulus updraft). Detrainment also applies to air that diverges from an updraft, for example, in the formation of anvil, which is likewise accompanied by mixing.

After mass dm of environmental air has been mixed into it, the parcel will have mass $m + dm$ and property $\mu' + d\mu'$. Conservation of μ then requires

$$(m + dm)(\mu' + d\mu') = m\mu' + dm\mu, \quad (7.16.1)$$

where $dm\mu$ represents how much of the conserved property has been incorporated into the parcel from the environment. If the above process occurs during time dt , expanding (7.16.1) and neglecting terms higher than first order leads to

$$\frac{d\mu'}{dt} = \frac{d(\ln m)}{dt}(\mu - \mu'), \quad (7.16.2)$$

which governs the conserved property inside the entraining updraft. More generally, if μ is not conserved but rather is produced per unit mass at the rate S_μ , its evolution inside the aforementioned parcel is governed by

$$\frac{d\mu'}{dt} = \frac{d(\ln m)}{dt}(\mu - \mu') + S_\mu. \quad (7.17)$$

The time rate of change following the parcel is related to its vertical velocity $w' = \frac{dz'}{dt}$ through the chain rule:

$$\frac{d}{dt} = w' \frac{d}{dz}, \quad (7.18)$$

where $\frac{d}{dz}$ refers to a change with respect to the parcel's height. Then (7.17) becomes

$$w' \frac{d\mu'}{dz} = \frac{w}{H_e}(\mu - \mu') + S_\mu, \quad (7.19)$$

in which $H_e^{-1} = \frac{d(\ln m)}{dz}$ defines the mass *entrainment height*. H_e represents the height for upwelling mass to increase by a factor of e . It characterizes the rate at which the entraining thermal expands due to incorporation of environmental air (Fig. 5.4).²

If $\mu = \ln \theta_e$, the source term $S_{\theta_e} = 0$ because θ_e is conserved. Then, with (5.27), (7.19) reduces to

$$\frac{d(\ln \theta'_e)}{dz} = \frac{1}{H_e} \left\{ \ln \left(\frac{\theta}{\theta'} \right) + \frac{l}{c_p} \left[\frac{r_c}{T} - \frac{r'_c}{T'} \right] \right\}. \quad (7.20)$$

Replacing saturation values by the actual mixing ratios and temperatures by those corresponding to the LCL (Sec. 5.4.2) and noting that differences of temperature are small compared with differences of moisture leads to

$$\begin{aligned} \frac{d(\ln \theta'_e)}{dz} &\cong \frac{1}{H_e} \left\{ \ln \left(\frac{\theta}{\theta'} \right) + \frac{l}{c_p T} (r - r') \right\} \\ &= \frac{1}{H_e} \left\{ \ln \left(\frac{T}{T'} \right) + \frac{l}{c_p T} (r - r') \right\}. \end{aligned} \quad (7.21)$$

Because $r' > r$ and $T' > T$, θ'_e decreases upward. This is to be contrasted with the parcel's behavior under conservative conditions (e.g., in the absence of entrainment).

² A cumulus cloud expands with height slower than the moist updraft supporting it because mixing with drier environmental air leads to evaporation of condensate and dissolution of cloud along its periphery (Sec. 9.3.3).

Its mass is then fixed and $\theta'_e = \text{const.}$ The two sinks of θ'_e on the right-hand side of (7.21) reflect transfers of sensible and latent heat to the environment. Those properties are mixed across the parcel's control surface (cf. Fig. 2.1).

Suppose now μ equals the specific momentum w . Then $w = 0$ for the environment. Equation (7.19) then reduces to

$$w \frac{dw}{dz} = -\frac{w^2}{H_e} + S_w$$

or

$$\frac{dK'}{dz} = -\frac{2}{H_e}K' + \frac{S_K}{w}, \quad (7.22)$$

where $K' = \frac{w^2}{2}$ is the specific kinetic energy of the parcel. The production of kinetic energy is related to the production of upward momentum: $S_K = wS_w$. Unlike the source term for θ_e , $S_K \neq 0$. K is not conserved, even in the absence of entrainment, because the parcel's kinetic energy changes through work that is performed on it by buoyancy.

Differentiating (7.12) along with (7.15) implies

$$\frac{dK'}{dz} = f_b \quad (7.23.1)$$

under conservative conditions (e.g., for a parcel of fixed mass, wherein $H_e \rightarrow \infty$). Hence we identify

$$\begin{aligned} S_K &= w f_b \\ &= wg \left(\frac{T' - T}{T} \right). \end{aligned} \quad (7.23.2)$$

Then (7.22) becomes

$$\frac{dK'}{dz} = g \left(\frac{T' - T}{T} \right) - \frac{2}{H_e}K'. \quad (7.24)$$

The sink of K' on the right hand side represents turbulent drag. It is exerted on the ascending parcel through the incorporation of momentum from the environment. By diluting kinetic energy, entrainment reduces the parcel's acceleration from what it would be under the action of buoyancy alone. That, in turn, limits its penetration into the stable layer above the LNB, to about an entrainment height (Prob. 7.21).

7.4.3 Potential instability

Until now, we have considered displacements of individual air parcels within a fixed layer. Suppose the layer itself is displaced vertically (e.g., in sloping convection or forced lifting over elevated terrain). Then changes of the layer's thermodynamic properties alter its stability. Stratification of moisture plays a key role in this process because different levels need not achieve saturation simultaneously.

Consider an unsaturated layer in which potential temperature increases with height,

$$\frac{d\theta}{dz} > 0, \quad (7.25.1)$$

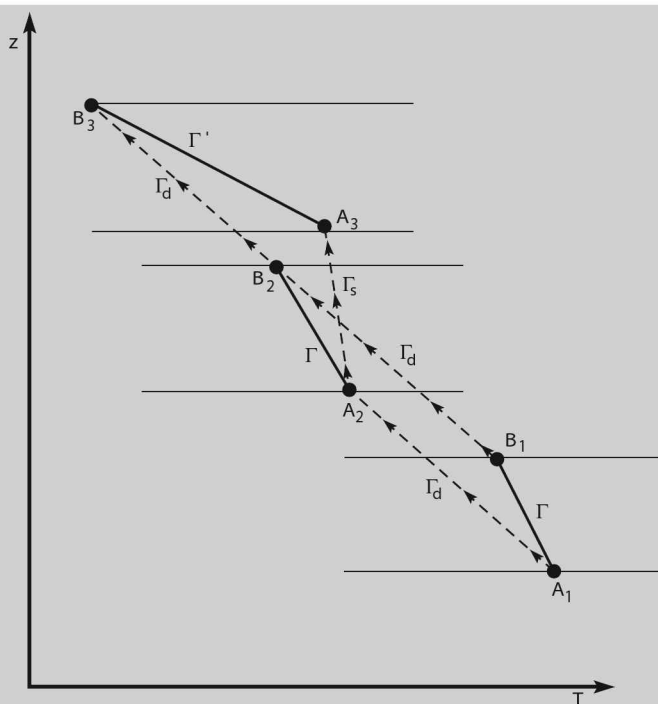


Figure 7.9 Successive positions and thermal structures of a potentially unstable layer (shaded), in which moisture decreases sharply with height, that is displaced vertically. Air is initially unsaturated, so all levels cool at the dry adiabatic lapse rate. However, displacement to position 2 leads to saturation of air at the layer's base (A_2). Thereafter, that air cools at the saturated adiabatic lapse rate, whereas air at the layer's top (B_2) continues to cool at the dry adiabatic lapse rate.

but in which equivalent potential temperature decreases with height,

$$\frac{d\theta_e}{dz} < 0. \quad (7.25.2)$$

By (7.8), the layer is stable. The difference $\theta_e - \theta$ for an individual parcel reflects the total latent heat available for release. Therefore, (7.25.2) represents a decrease with height of mixing ratio (e.g., as would develop over warm ocean).

Suppose that this layer is displaced vertically and that changes in its thickness can be ignored (Fig. 7.9). Because the layer is unsaturated, air parcels along a vertical section all cool at the dry adiabatic lapse rate Γ_d (e.g., between positions 1 and 2). The layer's profile of temperature is then preserved through the displacement. So is the layer's lapse rate and hence its stability.

Once some of the layer becomes saturated, this is no longer true. Lower levels, because they have greater mixing ratio, achieve saturation sooner than upper levels. Above their LCL, lower levels cool slower ($A_2 \rightarrow A_3$), at the saturated adiabatic lapse rate Γ_s . However, upper levels, which remain unsaturated, continue to cool at the dry

adiabatic lapse rate $\Gamma_d (B_2 \rightarrow B_3)$. Differential cooling between lower and upper levels then swings the temperature profile counter-clockwise, destabilizing the layer. This destabilization follows from the release of latent heat at lower levels, which heats the layer from below. Sufficient vertical displacement will render the layer unstable with respect to saturated conditions. This can occur through finite displacement, even if, initially, the layer was absolutely stable with respect to infinitesimal displacement. A layer for which this is possible is *potentially unstable*. The criterion for potential instability is (7.25.2). It implies an upward decrease of mixing ratio, as is often found above warm ocean.

A layer characterized by (7.25.2), but with the reversed inequality, is *potentially stable*. The criterion (7.25.2) then reflects an upward increase of mixing ratio. It accounts for the release of latent heat in upper levels, which, once they become saturated, stabilizes the layer. If a potentially stable or unstable layer becomes fully saturated, then it satisfies the usual criteria for instability - because all levels evolve in the same manner.

7.4.4 Modification of stability under unsaturated conditions

A layer's stability can also change through vertical compression and expansion, even under unsaturated conditions. Consider the layer in Fig. 7.10, which is displaced to a new height and thermodynamic state (distinguished by primed variables) from an initial height and state. Because θ is conserved under unsaturated conditions, the difference of θ across the layer is preserved:

$$d\theta' = d\theta.$$

Then the vertical gradient of potential temperature satisfies

$$\left(\frac{d\theta}{dz}\right)' dz' = \left(\frac{d\theta}{dz}\right) dz. \quad (7.26.1)$$

A horizontal section of layer having initial area dA must also satisfy conservation of mass:

$$\rho dA dz = \rho' dA' dz'. \quad (7.26.2)$$

Combining (7.26) obtains

$$\left(\frac{d\theta}{dz}\right)' = \left(\frac{d\theta}{dz}\right) \left(\frac{\rho'}{\rho}\right) \left(\frac{dA'}{dA}\right) \quad (7.27)$$

for the new vertical gradient of potential temperature inside the layer.

Sinking motion and horizontal divergence of air correspond to

$$\begin{aligned} \frac{\rho'}{\rho} &> 1 \\ \frac{dA'}{dA} &> 1, \end{aligned} \quad (7.28)$$

respectively. Each requires vertical compression of the layer: the former through a reduction of the layer's overall volume and the latter to compensate horizontal expansion of the layer. By (7.27), these motions steepen the vertical gradient of potential temperature. They increase the stability of a layer that is initially stable.

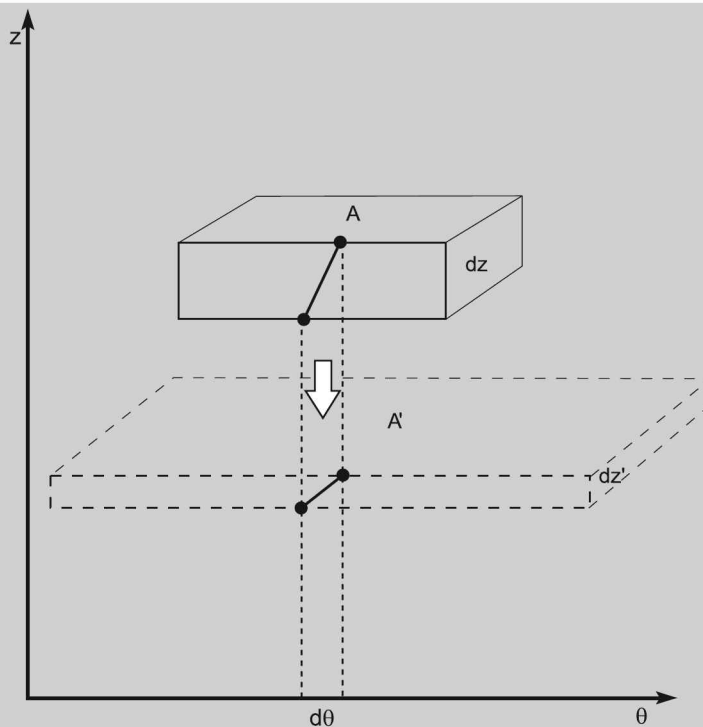


Figure 7.10 Successive positions and thermal structures of a layer that remains unsaturated and therefore preserves the potential temperature difference across it.

Similarly, they increase the instability of a layer that is initially unstable. For reasons developed in Chap. 12, stable air inside an anticyclone experiences sinking motion, or “subsidence.” It is accompanied near the surface by horizontal divergence. Together, these features of motion inside an anticyclone intensify stability. They lead to the formation of a *subsidence inversion*. Found at heights as low as a kilometer, it caps vertical dispersion, trapping pollutants near the ground.

In a cyclone, just the reverse occurs: Air experiences ascending motion and horizontal convergence. These features of motion inside a cyclone imply (7.28) with the inequalities reversed. The direction in Fig. 7.10 is likewise reversed. Vertical expansion then weakens the vertical gradient of potential temperature. It drives a stable layer towards neutral stability. Ascending motion and horizontal convergence inside a cyclone can eliminate the stability of a layer, providing conditions favorable for convection.

7.5 STABILIZING AND DESTABILIZING INFLUENCES

The preceding development illustrates that stability can be modified by internal motion that rearranges air. More important to the overall stability of the atmosphere is external heat transfer, which shapes thermal structure. The vertical distribution of θ provides

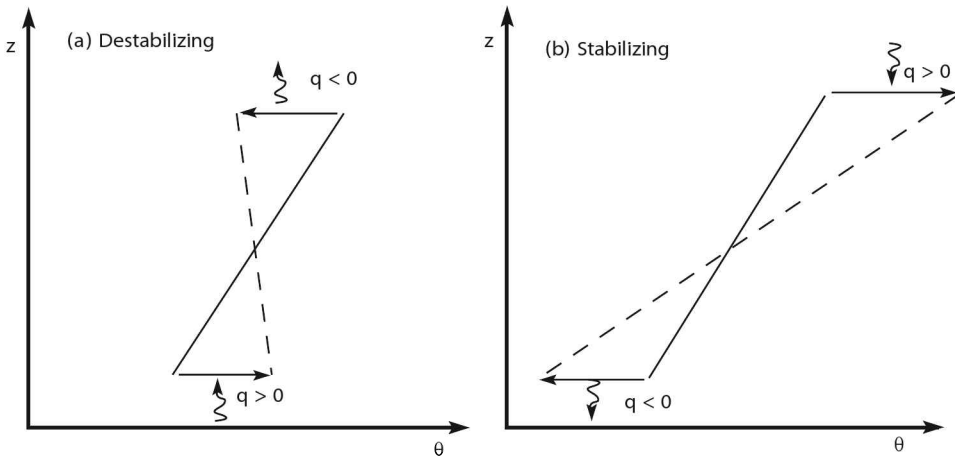


Figure 7.11 (a) Destabilizing influences: Heating from below and cooling from above act to swing the profile of potential temperature counterclockwise. (b) Stabilizing influences: Cooling from below and heating from above act to swing the profile of potential temperature clockwise.

insight into how stability is established because it reflects the effective stratification of mass.

In light of the First Law (2.36), the stability criteria (7.8) imply that a layer is destabilized by heating from below or cooling from aloft (Fig. 7.11a). Each rotates the vertical profile of θ counter-clockwise and thus toward instability. These are precisely the influences exerted on the troposphere by exchanges of energy with the Earth's surface and deep space. Absorption of LW radiation and transfers of latent and sensible heat from the Earth's surface warm the troposphere from below. LW emission at upper levels cools the troposphere from above. Both drive the vertical profile of θ counter-clockwise and the lapse rate toward superadiabatic values. By destabilizing the stratification, these forms of heat transfer maintain convective overturning. The latter, because it operates on short time scales, maintains the troposphere close to moist neutral stability. It makes the general circulation of the troposphere behave as a heat engine (Fig. 6.7), driven thermally by heat transfer from the Earth's surface.

Conversely, cooling from below or heating from above stabilize a layer (Fig. 7.11b). Each rotates the profile of θ clockwise and thus toward increased stability. These forms of heat transfer are symbolic of the stratosphere. There, ozone heating increases upward, stabilizing thermal structure. By inhibiting vertical motion, strong stability suppresses turbulent dispersion, forcing constituents to become stratified. Work must then be performed against buoyancy to drive vertical motion. This makes the general circulation of the stratosphere behave as a refrigerator (Fig. 6.8), driven mechanically by the absorption of wave activity that propagates up from below.

According to Secs. 7.2 and 7.4, moisture also figures in the stability of a layer. If the lapse rate is conditionally unstable (7.9.3), humidifying a layer lowers the LCL and hence the LFC of individual parcels. That, in turn, increases CAPE, which fuels deep convection (cf. Fig. 7.6). Likewise, humidifying lower portions of the layer (e.g., through contact with warm ocean) increases the equivalent potential temperature there. This

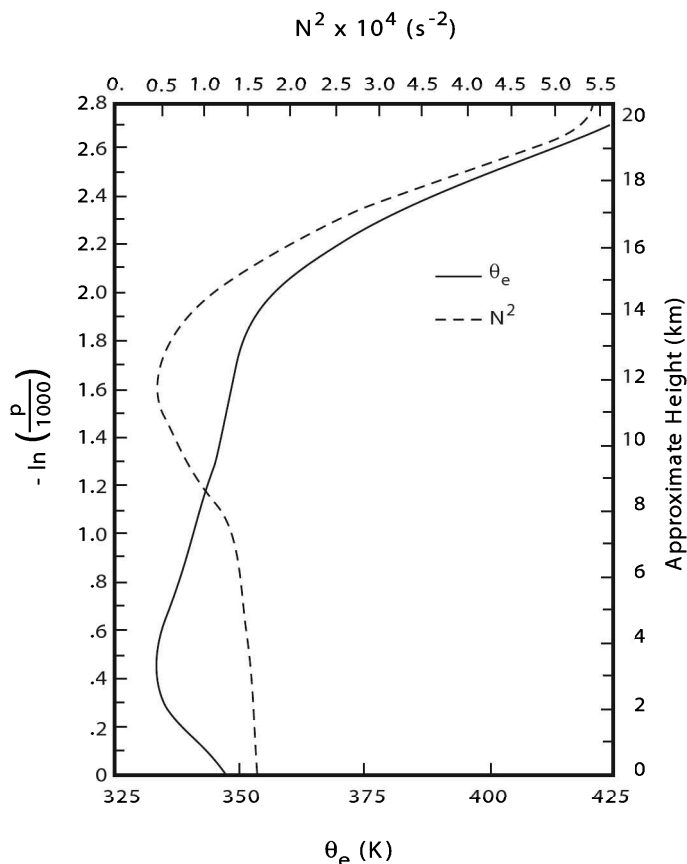


Figure 7.12 Mean vertical profiles of equivalent potential temperature (solid) and buoyancy frequency squared (dashed) over Indonesia, obtained by averaging radiosonde observations between longitudes of 90E and 105E and latitudes of 10S and 10N during March 1984 (cf. Fig. 1.30).

reflects a transfer of latent heat across the layer's lower boundary. It swings the profile of θ_e counter-clockwise, driving $\frac{d\theta_e}{dz}$ toward negative values. Thus, introducing moisture from below, as occurs over warm ocean, drives the troposphere toward potential instability.

For conditional and potential instability, finite displacements of air - whether they be of an individual air parcel or of a layer *in toto* - eventually result in buoyantly driven convection. However, for each, a finite potential well must be overcome before the instability can be released. Due to the source of water vapor at its base, lower portions of the tropical troposphere are potentially unstable. Figure 7.12 presents mean distributions of θ_e and N^2 over the equatorial western Pacific. $N^2 > 0$ implies that θ increases upward throughout (7.10.3). The layer is stable with respect to unsaturated conditions. However, in the lowest 3 km, θ_e decreases upward, making the layer potentially unstable. Despite such instability and, more generally, conditional instability, deep convection develops only in preferred regions, those where forced lifting is

prevalent (Fig. 1.30). Over the Indian Ocean and western Pacific, equatorward-moving air is driven by thermal contrast between ocean and neighboring continent. It converges horizontally to produce forced ascent. The latter lifts air above the LFC, releasing it from the potential well that constrains air at lower levels. Similarly, diurnal heating of tropical landmasses imparts enough buoyancy for surface air over the maritime continent, South America, and tropical Africa to likewise overcome the potential well. Once released at the LFC, that air accelerates upward through work that is performed by buoyancy.

7.6 TURBULENT DISPERSION

7.6.1 Convective mixing

Vertical stability controls the development of vertical motion and hence mixing by 3D turbulence. Regions of weak or negative stability favor convective overturning, which results in efficient mixing of air. This situation is common near the ground because absorption of SW radiation destabilizes the surface layer. Introducing moisture from below has a similar effect. By increasing the equivalent potential temperature of surface air, it makes the overlying layer potentially unstable. When conditions are favorable, small disturbances evolve into fully developed convection. It, in turn, neutralizes the instability by rearranging mass vertically.

As noted earlier, the troposphere has a global-mean lapse rate of about 6.5 K km^{-1} . This is close to neutral stability with respect to saturated conditions ($\Gamma_s \cong 5.5 \text{ }^\circ\text{K/km}$). The mean stratification reflects two salient features of the troposphere: (1) Tropospheric air remains moisture laden, through contact with warm ocean. (2) It is efficiently overturned in cumulus and sloping convection. Efficient vertical exchange carries air between the surface and the upper troposphere in a matter of days, only hours inside cumulus towers. Because it operates on a time scale that is short compared with diabatic influences, convective mixing drives thermal structure of the troposphere toward $\theta_e = \text{const}$ and a mean state of moist neutral stability (Sec. 6.5).

Cumulus convection is favored in the tropical troposphere by high SST and strong insolation. Both support large transfers of radiative, latent, and sensible heat from the Earth's surface, which destabilize the troposphere. Opposing those destabilizing influences is cumulus convection. Reinforced by the release of latent heat, it maintains much of the tropical troposphere close to moist neutral stability. Cumulus convection also operates outside the tropics. However, there, it is organized by synoptic weather systems. Frontal motions associated with cyclones, like the one in the eastern Atlantic in Fig. 1.28, lift air through sloping convection. By displacing moist air that originated in the tropics, often conditionally or potentially unstable, those large-scale motions favor cumulus convection. Hence, they too drive stratification toward moist neutral stability.

7.6.2 Inversions

Contrary to the troposphere, the stratosphere is characterized by strong positive stability. Ozone heating makes temperature increase upward. Potential temperature then increases upward sharply (Fig. 6.5), from about 350 K near the tropopause to more than 1000 K near 10 hPa. The upward vertical gradient of θ produces a strong

positive restoring force. It suppresses vertical motion and 3D turbulence, forcing chemical constituents to become stratified.

Unlike the troposphere, where air is efficiently overturned, the stratosphere has a characteristic time scale for vertical exchange of months to years. Its layered structure is occasionally illustrated by volcanic eruptions, which hurl debris above the tropopause. During the 1950s and 1960s, nuclear tests provided similar illustrations. Aerosol particles introduced by those sources have very slow settling rates. Consequently, they move with individual bodies of air, behaving as tracers. Observed by satellite, stratospheric aerosol exhibits little or no vertical motion. Instead, it fans out horizontally, forming cloud of global dimension that remains in the stratosphere for many months (see Fig. 9.6).

The long residence time of stratospheric aerosol enables it to alter SW absorption at the Earth's surface for extended durations. This feature makes volcanic aerosol a consideration for climate. The great eruption of Krakatoa destroyed the Indonesian island of the same name in 1883. Manifest in the color of sunsets, it modified radiative properties of the stratosphere for years afterwards. Pinatubo exerted a similar influence. As is apparent in Fig. 1.27, Pinatubo influenced global-mean surface temperature by reducing the transmission of SW radiation. Through cumulus convection, which was affected indirectly, it also influenced water vapor in the upper troposphere.

Tropospheric aerosol is introduced routinely by deforestation. It was magnified during the Kuwaiti oil fires, which followed the Persian Gulf War. In contrast to stratospheric aerosol, tropospheric aerosol is not long lived. It is removed on a time scale of only days, by rain out and other scavenging processes associated with convection (Chap. 9).

Convection does not penetrate appreciably above the tropopause, where N^2 and the restoring force of buoyancy increase sharply (Fig. 7.12). Upon encountering the sharp increase of stability, a cumulus updraft is quickly drained of positive buoyancy. It fans out horizontally to form cloud anvil, as typifies the mature stage of cumulonimbus thunderstorms (Fig 7.7). Similar behavior is occasionally observed at intermediate levels of the troposphere, when amplifying cumulus encounter a shallow inversion through which the updraft is able to penetrate (see Fig. 9.18).

Temperature structure inside an anticyclone is often inverted. A subsidence inversion can form less than a kilometer above the surface (Sec. 7.4). It confines pollutants to a shallow layer near the ground, where they are introduced. Because wind inside an anticyclone also tends to be light, pollutants are neither dispersed nor advected away from source regions. Their concentrations therefore increase during periods of high emission. The same mechanisms inhibit cloud formation by confining moisture near the surface. Consequently, anticyclones are typified by cloud-free but polluted conditions.

7.6.3 Life cycle of the nocturnal inversion

Heat transfer from the Earth's surface and convection compete for control of stratification. Their competition is illustrated by the formation and breakup of the *nocturnal inversion*, strong stability near the surface that often develops at night. Arid and cloud-free conditions permit efficient LW cooling to space. Surface temperature then decreases sharply after sunset. To preserve thermal equilibrium, heat is absorbed, radiatively and conductively, from the overlying air. This process cools the surface

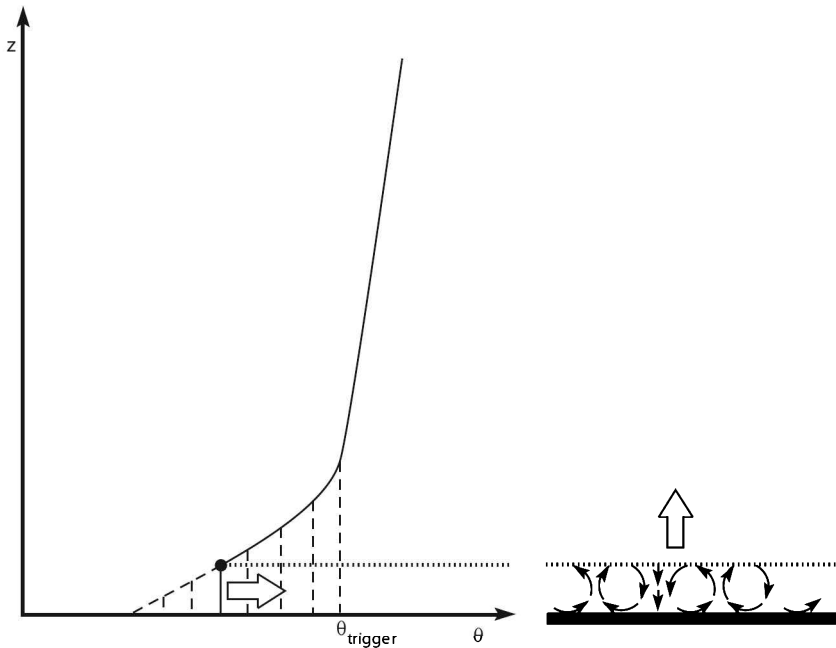


Figure 7.13 Schematic illustrating the breakdown of the nocturnal inversion, which forms during night. Following sunrise, surface heating and heat transfer to the overlying air destabilize a shallow layer adjacent to the ground. Convection then develops spontaneously and, through vertical mixing, restores the thermal structure inside that layer to neutral stability. Continued heat transfer from the ground requires θ to be mixed over a progressively deeper layer, which erodes the nocturnal inversion from below. When convection has advanced through the entire layer of strong stability, the nocturnal inversion has been destroyed. Deep convection can then penetrate into weakly stable air overhead, dispersing pollutants that were previously trapped near the surface in stable air.

layer from below, stabilizing it. Sufficient cooling will swing the temperature profile inverted (Fig. 7.13). Potential temperature near the ground then increases upward sharply.

Strongly stable, the nocturnal inversion suppresses vertical motion. It also suppresses 3D turbulence, which requires vertical eddy motion. Without turbulent dispersion, pollutants emitted at the surface are frozen in the stable air. Their concentrations therefore increase steadily near pollution sources. Such behavior is particularly noticeable in urban areas, during early morning periods of heavy traffic.

After sunrise, absorption of SW radiation warms the ground. It influences the surface layer in just the reverse manner. To preserve thermal equilibrium, heat is then transferred to the overlying air, initially through conduction and radiative transfer. These forms of heat transfer warm the surface layer from below. They increase potential temperature near the ground, swinging its vertical profile counter clockwise. In this manner, a shallow layer near the surface is destabilized. Thermal convection then develops spontaneously, neutralizing the instability. By rearranging air vertically, heat

is transferred from the surface (high θ) and mixed vertically across a shallow layer of convective overturning. Inside that layer, θ is homogenized, neutralizing the layer's stability. The same process operates on other conserved properties, such as the mixing ratios of water vapor and pollutants. Like θ , each is driven toward a uniform vertical profile inside the layer of convective overturning (see Fig 13.6).

Absorption of SW radiation and upward heat transfer continue to destabilize the surface layer. That, in turn, maintains convective overturning. To preserve neutral stability, additional heat transferred to the atmosphere must be mixed vertically over a deepening layer. The layer driven toward neutral stability ($\theta \cong \text{const}$) therefore expands upward. So do pollutants that were frozen near the surface in stable air. As surface heating continues, convective mixing systematically erodes the nocturnal inversion from below (Fig. 7.13). Simultaneously, it reduces pollution concentrations by diluting them over an ever deeper layer.

After sufficient surface heating, convection will have advanced through the entire nocturnal inversion. At that point, it reaches air of weak stability overhead (e.g., air that was unaffected by surface cooling during the night). The surface temperature at which this condition is achieved is the *break* or *trigger temperature*. At this temperature, deep convection can develop from the source of positive buoyancy at the ground. Pollutants are then dispersed vertically over a deep volume. They are also dispersed horizontally by 3D turbulence, which develops spontaneously. Together, these forms of transport result in a sharp reduction of pollution concentrations. The latter is often attended by an abrupt improvement in visibility. By permitting vertical transport of water vapor, the breakup of the nocturnal inversion also enables the formation of convective cloud. Cumulus appears soon thereafter.

7.7 RELATIONSHIP TO OBSERVED THERMAL STRUCTURE

Thermal structure is strongly influenced by convection, which operates on a time scale much shorter than radiative heating. In the tropics, cumulus convection is the dominant form of vertical motion. The height to which it prevails is dictated by CAPE (Sec. 7.4.1). Produced by transfers of heat from the Earth's surface, CAPE determines the LNB, which caps deep convection. It is reflected in the θ_e of surface air.

An air parcel initially at the surface will, if lifted above the LFC, accelerate upward. By the time it reaches the upper troposphere, nearly all of the parcel's water vapor will have been removed through condensation, attended by the release of its latent heat. The pseudo isentrope ($\theta'_e = \text{const}$) along which the parcel evolves will then coincide with an isentrope ($\theta' = \text{const}$); see Prob. 5.20. Because $p' = p$ (mechanical equilibrium), the condition of neutral buoyancy that defines the LNB (7.3)

$$T'_{LNB} = T_{LNB} \quad (7.29.1)$$

is then equivalent to

$$\theta'_{LNB} = \theta_{LNB}. \quad (7.29.2)$$

As the parcel is now devoid of water vapor, this reduces to the condition

$$\theta_{LNB} \cong \theta'_e. \quad (7.30)$$

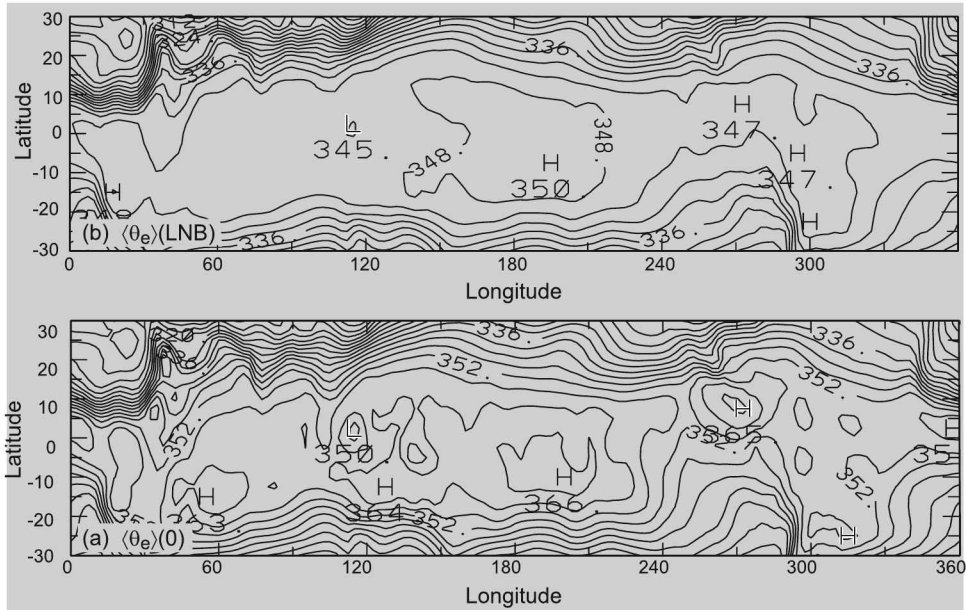


Figure 7.14 (a) Geographical distribution of equivalent potential temperature at the Earth's surface (mean plus 1 std deviation), $\theta_e(0)$. (b) As in (a), but at the Level of Neutral Buoyancy (LNB), accounting for entrainment of cumulus updrafts (7.21). After Salby et al. (2003).

The LNB corresponds to the height where environmental θ equals θ'_e of cumulus updrafts.

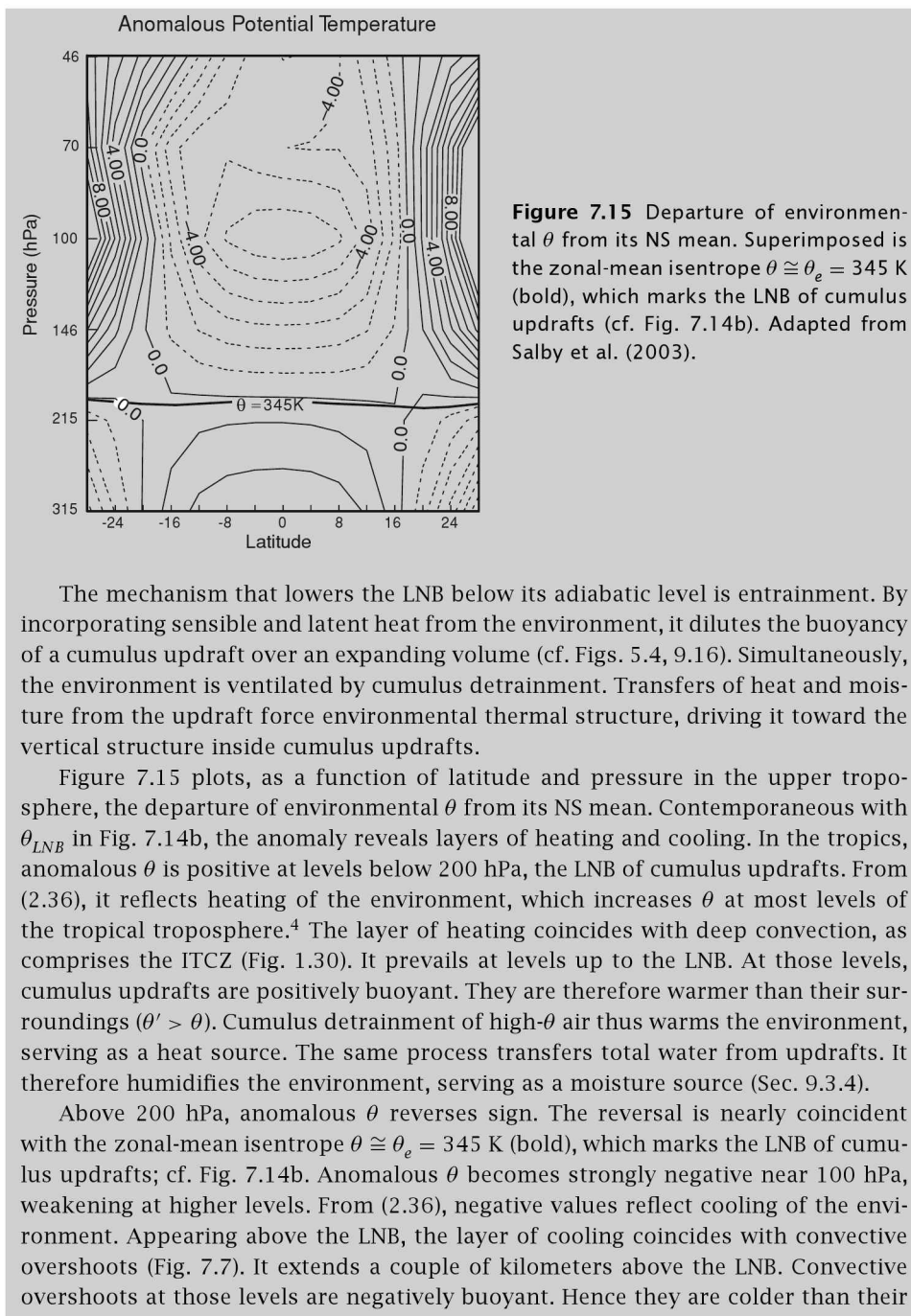
Under adiabatic conditions (e.g., in the absence of entrainment), (7.30) simplifies to

$$\theta_{LNB} = \theta_e(z = 0)$$

because θ'_e is then conserved, equal to its initial value. In the presence of entrainment, θ'_e decreases upward. Reflecting diabatic losses of sensible and latent heat, this reduces the limiting values of θ' , which is no longer constant even in the upper troposphere. By (7.29.2), entrainment therefore lowers the LNB.

Plotted in Fig. 7.14a is the horizontal distribution of $\theta_e(z = 0)$ during November 1991 through February 1992. Values maximize in the equatorial Pacific, where $\theta_e(z = 0)$ approaches 370 K. That value corresponds to θ_{LNB} under adiabatic conditions. The isentropic surface with $\theta = 370$ K is positioned near the tropical tropopause, just above 100 hPa. Although deep cumulus form in the equatorial Pacific during this period, few reach this high. Instead, cloud coverage decreases sharply above 200 hPa (see Figs 9.24, 9.25).

Plotted in Fig. 7.14b is the horizontal distribution of θ_{LNB} , with entrainment accounted for through (7.21). The pattern resembles that of $\theta_e(z = 0)$, but with values reduced. θ_{LNB} has values of about 345 K over most of the tropics, 350 K in the equatorial Pacific. $\theta = 345$ K positions the LNB near 200 hPa, several kilometers lower than is possible under adiabatic conditions. It is at this level that cloud anvil is most extensive and air diverges horizontally (Fig. 9.26a).



The mechanism that lowers the LNB below its adiabatic level is entrainment. By incorporating sensible and latent heat from the environment, it dilutes the buoyancy of a cumulus updraft over an expanding volume (cf. Figs. 5.4, 9.16). Simultaneously, the environment is ventilated by cumulus detrainment. Transfers of heat and moisture from the updraft force environmental thermal structure, driving it toward the vertical structure inside cumulus updrafts.

Figure 7.15 plots, as a function of latitude and pressure in the upper troposphere, the departure of environmental θ from its NS mean. Contemporaneous with θ_{LNB} in Fig. 7.14b, the anomaly reveals layers of heating and cooling. In the tropics, anomalous θ is positive at levels below 200 hPa, the LNB of cumulus updrafts. From (2.36), it reflects heating of the environment, which increases θ at most levels of the tropical troposphere.⁴ The layer of heating coincides with deep convection, as comprises the ITCZ (Fig. 1.30). It prevails at levels up to the LNB. At those levels, cumulus updrafts are positively buoyant. They are therefore warmer than their surroundings ($\theta' > \theta$). Cumulus detrainment of high- θ air thus warms the environment, serving as a heat source. The same process transfers total water from updrafts. It therefore humidifies the environment, serving as a moisture source (Sec. 9.3.4).

Above 200 hPa, anomalous θ reverses sign. The reversal is nearly coincident with the zonal-mean isentrope $\theta \cong \theta_e = 345$ K (bold), which marks the LNB of cumulus updrafts; cf. Fig. 7.14b. Anomalous θ becomes strongly negative near 100 hPa, weakening at higher levels. From (2.36), negative values reflect cooling of the environment. Appearing above the LNB, the layer of cooling coincides with convective overshoots (Fig. 7.7). It extends a couple of kilometers above the LNB. Convective overshoots at those levels are negatively buoyant. Hence they are colder than their

⁴ Except at lower levels, areal coverage by cloud is small. It decreases in the upper troposphere to only a few percent (cf. Fig. 9.24). Consequently, large-scale thermal structure in Fig. 7.15 describes regions that are chiefly cloud-free.

surroundings ($\theta' < \theta$). Cumulus detrainment of low- θ air therefore cools the environment, serving as a heat sink. This process drives environmental thermal structure toward the nearly uniform vertical distribution of θ inside cumulus updrafts. By driving Γ toward Γ_d and neutralizing stability, it acts to raise the tropopause and decrease its temperature (cf. Fig. 7.8).

SUGGESTED REFERENCES

Vertical stability is treated in detail in *Atmospheric Thermodynamics* (1981) by Iribarne and Godson.

Environmental Aerodynamics (1978) by Scorer contains an advanced treatment of entrainment, including laboratory simulations.

PROBLEMS

A pseudo-adiabatic chart is provided in Appendix F.

1. Use Newton's 2nd law to derive the momentum balance (7.1.1) for an individual air parcel.
2. In early morning, the profile of potential temperature in a layer adjacent to the surface is described by

$$\frac{\theta}{\theta_0} = e^{\frac{z}{h_1}} - \frac{z}{h_2},$$

with $h_1 > h_2$. If condensation can be ignored, (a) characterize the static stability of this layer, (b) determine the profile of buoyancy frequency, (c) determine the limiting distribution of θ for $t \rightarrow \infty$, after sufficient surface heating for convection to develop.

3. Use Archimedes' principle to establish (7.2.2).
4. A simple harmonic oscillator comprised of a 1-kg weight and an elastic spring with constant k provides an analogue of buoyancy oscillations. Calculate the effective spring constant for unsaturated conditions representative of (a) the troposphere: $\Gamma = 6.5 \text{ }^\circ\text{K km}^{-1}$ and $T = 260 \text{ }^\circ\text{K}$, (b) the stratosphere: $\Gamma = -4 \text{ }^\circ\text{K km}^{-1}$ and $T = 250 \text{ }^\circ\text{K}$.
5. An impulsive disturbance imposes a vertical velocity w_0 on an individual air parcel. If the layer containing that parcel has nonnegative stability N , (a) determine the parcel's maximum vertical displacement, as a function of N , under linear conditions, (b) describe the layer's evolution under nonlinear conditions, in relation to N .
6. Upstream of isolated rough terrain, the surface layer has thermal structure

$$T(z) = T_0 \frac{a}{a+z},$$

where $a > 0$ and $T_0 > a\Gamma_d$. (a) Determine the upstream profile of potential temperature. (b) Characterize the upstream vertical stability. (c) Describe the mean stratification far downstream of the rough terrain.

7. Derive the identity

$$\frac{1}{\theta} \frac{d\theta}{dz} = \frac{1}{T} \frac{dT}{dz} + \frac{\kappa}{H}$$

which relates the vertical gradients of temperature and potential temperature.

8. An early morning sounding reveals the temperature profile

$$T(z) = \begin{cases} T_0 + \left(\frac{\Gamma_d}{2}\right)z & z < 1 \text{ km} \\ T_0 + \left(\frac{\Gamma_d}{2}\right) - \left(\frac{\Gamma_d}{4}\right)(z - 1) & z \geq 1 \text{ km}. \end{cases}$$

(a) Determine the *trigger temperature* for deep convection to develop. (b) If, following sunrise, the ground warms at 3°C per hour, estimate the time when cumulus clouds will appear.

9. Radiative heating leads to temperature in the stratosphere increasing up to the stratopause, above which it decreases in the mesosphere. If the temperature profile at a certain station is approximated by

$$T(z) = T_0 \frac{z}{b + \left(\frac{z}{h}\right)^2},$$

determine (a) the profile of potential temperature, (b) the profile of Brunt-Väisällä frequency squared N^2 . (c) For reasons developed in Chap. 14, gravity waves can propagate through regions of positive static stability, which provides the positive restoring force for their oscillations. In terms of the variable b , how small must h be to block propagation through the mesosphere? (d) By noting that gravity waves introduce vertical displacements, discuss the behavior of air in the mesosphere under the conditions in (c).

10. Analyze the stability and energetics for the stratification and parcel in Fig. 7.6, but under initially saturated conditions.
11. (a) To what height would the parcel in Prob. 7.10 penetrate under purely adiabatic conditions? (b) How/why does this differ from reality? (See Prob. 7.21)
12. Estimate the maximum height of convective towers under the conditions in Fig. 7.6, but with the isothermal stratosphere replaced by one with a lapse rate of -3 °K km^{-1} .
13. The approach of a warm front is often heralded by deteriorating visibility and air quality. Consider a frontal surface that slopes upward and to the right, separates warm air on the left from cold air on the right, and moves to the right. The intersection of this surface with the ground marks the front. (a) Characterize the vertical stability at a station ahead of the front. (b) Discuss the vertical dispersion of pollutants that are introduced at the ground. (c) Describe how the features in (a) and (b) and pollution concentrations evolve as the front approaches the station.
14. A morning temperature sounding over Florida reveals the profile

$$T = \begin{cases} 20 - 8(z - 1) \text{ (°C)} & z \geq 1 \text{ km} \\ 20 \text{ (°C)} & z < 1 \text{ km}. \end{cases}$$

If the mixing ratio at the surface is 20 g kg^{-1} , determine the LFC.

15. The potential temperature inside a layer varies as

$$\frac{\theta}{\theta_0} = -(z - a)^2 + a^2 + 1 \quad z > 0.$$

(a) Characterize the stability of this layer. (b) How large a vertical displacement would a parcel originally at $z = 0$ have to undergo to be released from this layer?

16. Establish stability criteria analogous to (7.8) under saturated conditions.

17. Rayleigh friction approximates turbulent drag in proportion to an air parcel's velocity:

$$D = Kw,$$

where D is the specific drag experienced by the parcel, w is its velocity, and the Rayleigh friction coefficient K has dimensions of inverse time. Show that, in the presence of Rayleigh friction, vertical displacements remain bounded under conditions of neutral stability.

18. Show that, for given Rayleigh friction coefficient K , vertical displacements remain bounded for small instability but become unbounded for sufficiently large instability.
19. The lapse rate at a certain tropical station is constant and equal to 7°K km^{-1} from 1000 hPa to 200 hPa, above which conditions are isothermal. At the surface, the temperature is 30°C and the relative humidity is sufficiently great for vertical displacements to achieve immediate saturation. Approximate the saturated adiabat by the constant lapse rate: $\Gamma_s \cong 6.5^\circ\text{K km}^{-1}$, to estimate the maximum vertical velocity attainable inside convective towers. (see Prob. 7.21)
20. The weather service forecasts the intensity of convection in terms of a *thermal index TI*, which is defined as the local ambient temperature minus the temperature anticipated inside a thermal under adiabatic conditions. The more negative is TI , the stronger are updrafts inside thermals. A morning sounding over an arid region reveals a lapse rate of -4°K km^{-1} in the lowest kilometer and a constant lapse rate of 6°K km^{-1} above, with the surface temperature being 10°C . If air is sufficiently dry to remain unsaturated and if the surface temperature is forecast to reach a maximum of 40°C , (a) determine the thermal index then as a function of height for $z > 1$ km, (b) to what height will updrafts intensify? (c) Calculate the maximum updraft and maximum height reached by updrafts anticipated under adiabatic conditions if, at the time of maximum temperature, mean thermal structure has been driven adiabatic in the lowest kilometer.
21. In practice, vertical velocities inside a thermal differ significantly from adiabatic values. Entrainment limits the maximum updraft to that near the level of maximum buoyancy. Under the conditions in Prob. 7.20, but accounting for entrainment, calculate the profile of kinetic energy and note the height of maximum updraft inside a thermal for an entrainment height of (a) infinity, (b) 1 km, (c) 100 m.
22. Establish relationship (7.21) that governs behavior inside an entraining thermal.
23. Consider the conditions in Prob. 7.19, but in light of entrainment. If the ambient mixing ratio varies as $r = r_0 e^{-z/h}$, with $h = 2$ km, and the entrainment height is 1 km, calculate the profile of equivalent potential temperature inside a moist thermal.
24. Consider a constant tropospheric lapse rate of $\Gamma = 7^\circ\text{K km}^{-1}$ and mixing ratio that varies as $r = r_0 e^{-z/h}$, with $h = 2$ km. If the troposphere can be regarded as a layer close to saturation, evaluate its stability to large-scale lifting (e.g., in sloping convection) in the presence of moisture that is representative of (a) mid-latitudes: $r_0 = 5 \text{ g kg}^{-1}$, (b) the tropics: $r_0 = 25 \text{ g kg}^{-1}$.

Radiative transfer

The thermal structure and stratification discussed in Chaps. 6 and 7 are shaped in large part by radiative transfer. According to the global-mean energy budget (Fig. 1.32), of some 542 W m^{-2} that is absorbed by the atmosphere, more than 80% is supplied through radiative transfer: 368 W m^{-2} through absorption of LW radiation from the Earth's surface and another 68 W m^{-2} through direct absorption of SW radiation. Similarly, the atmosphere loses energy through LW emission to the Earth's surface and to space.

To maintain thermal equilibrium, components of the atmospheric energy budget must, on average, balance. Vertical transfers of radiant energy involved in this balance are instrumental in determining thermal structure and motion that characterize individual layers. Likewise, the dependence of radiative transfer on latitude, temperature, and cloud introduces horizontal variations into the energy budget. Those variations lead to net radiative heating at low latitude and net radiative cooling at middle and high latitudes (Fig. 1.34c). To maintain thermal equilibrium locally, the latter must be compensated by a poleward transfer of heat. This heat transfer is accomplished by the general circulation. The general circulation, in turn, is driven by the horizontal distribution of radiative heating and cooling. Understanding these features of the Earth-atmosphere system therefore requires an understanding of how radiation interacts with the atmosphere.

8.1 SHORTWAVE AND LONGWAVE RADIATION

Energy transfer in the atmosphere involves radiation in two distinct bands of wavelength: shortwave radiation emitted by the sun and longwave radiation emitted by the Earth's surface and atmosphere. Owing to the disparate temperatures of the emitters, these two forms of radiation are concentrated at wavelengths λ that are widely

separated. Figure 8.1a shows blackbody emission spectra for temperatures of 6000°K and 288°K, corresponding to the equivalent blackbody temperature of the sun and the global-mean surface temperature of the Earth.¹ The spectrum of SW radiation is concentrated at wavelengths shorter than 4 μm . It peaks in the visible near $\lambda = 0.5 \mu\text{m}$. Wings of the SW spectrum extend into the UV ($\lambda < 0.3 \mu\text{m}$) and into the near-infrared ($\lambda > 0.7 \mu\text{m}$). On the other hand, the spectrum of LW radiation peaks well into the IR, at a wavelength of about 10 μm . Wings of the LW spectrum extend down to wavelengths of about 5 μm and out to the microwave region ($\lambda > 100 \mu\text{m}$). Overlap between the spectra of SW and LW radiation is negligible. They may therefore be treated independently by considering wavelengths shorter and longer than 4 μm .

Also shown as a function of wavelength is the fractional absorption of radiation passing from the top of the atmosphere (TOA) to a height of 11 km (Fig. 8.1b) and to the ground (Fig. 8.1c). Energetic radiation: $\lambda < 0.3 \mu\text{m}$, is absorbed at high levels through photodissociation and ionization of O_2 and O_3 . As a result, very little of the UV incident on TOA reaches the tropopause. By contrast, most of the spectrum in the visible and near-IR arrives at the tropopause unattenuated. However, in passing from 11 km to the ground, the remaining SW spectrum is substantially absorbed in the IR by water vapor and carbon dioxide, absolute concentrations of which are large in the troposphere. Consequently, the spectrum of SW radiation reaching the surface is concentrated at visible wavelengths. To those wavelengths, the atmosphere is mostly transparent.

In contrast to SW radiation, LW radiation emitted by the Earth's surface is almost completely absorbed - by H_2O across a wide band centered at 6.3 μm and another in the far-infrared, by CO_2 in a band centered at 15 μm (near the peak of the LW emission spectrum), and by a variety of trace gases including O_3 , CH_4 , and N_2O . Therefore, most of the LW energy emitted by the ground is captured in the overlying layer of air. That energy must be re-emitted, half upward and half back downward. The upward component is then absorbed again in the next overlying layer. It must likewise be re-emitted, half upward and half back downward. In this fashion, energy emitted by the Earth's surface undergoes repeated absorption and re-emission in adjacent atmospheric layers until, eventually, it is rejected to space. This sequence of radiative exchanges traps LW radiation and, through mechanisms developed below, elevates the surface temperature of the Earth.

Comparing Figs. 8.1b and 8.1c indicates that most of the LW energy emitted by the Earth's surface is absorbed in the troposphere. Only in the *atmospheric window* at wavelengths of 8-12 μm is absorption weak enough for much of the LW radiation emitted by the surface to pass freely to space. The 9.6- μm band of ozone, which is positioned inside this window, is the only strong absorber at those wavelengths. Most of that absorption takes place in the stratosphere, where ozone is concentrated.

8.1.1 Spectra of observed SW and LW radiation

The most important property of the SW spectrum is the *solar constant*, which equals the flux of radiant energy that arrives at TOA (at the mean Earth-sun distance), integrated over wavelength. The solar constant is weakly variable, having a value of about

¹ The representation in Fig. 8.1a, logarithmic in wavelength versus wavelength times energy flux, is area preserving. Equal areas under a curve in different bands of wavelength represent equal power.

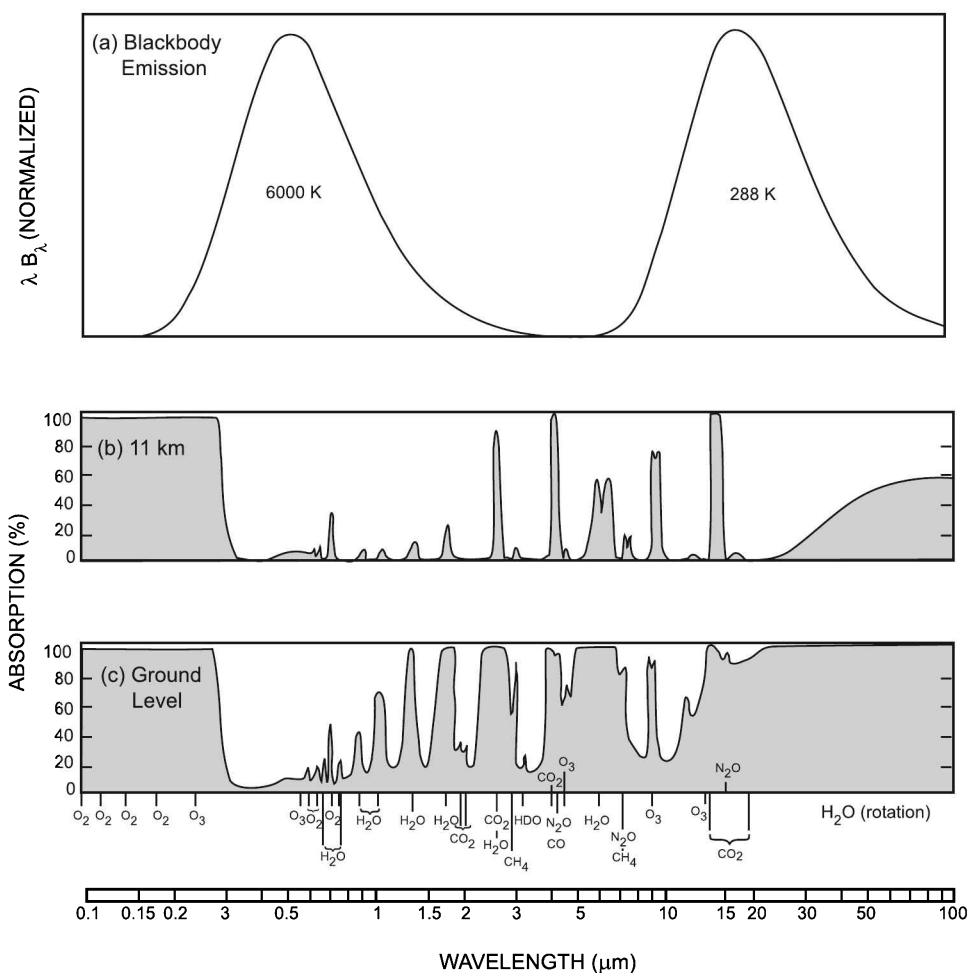


Figure 8.1 (a) Blackbody emission spectra as functions of wavelength λ for temperatures corresponding to the sun and the Earth's global-mean surface temperature. The representation: λ times energy flux vs $\log(\lambda)$, is area preserving – equal areas under a spectral curve represent equal power. (b) Absorption as a function of λ along a vertical path from the top of the atmosphere to a height of 11 km. (c) As in (b), but to the surface. Adapted from Goody and Yung (1989).

1370 W m^{-2} . When distributed over the Earth, this SW flux leads to the daily *insolation* shown in Fig. 1.33, which follows primarily from the length of day and solar inclination for particular latitudes and times of year.

Spectra of observed SW and LW radiation are shaped by the absorption characteristics described above. At TOA (Fig. 8.2), the solar spectrum resembles the emission spectrum of a blackbody at about 6000°K . That temperature is characteristic of the sun's *photosphere*, where most of the radiation is emitted. Wavelengths shorter than 300 nm deviate from this simple picture. Instead, they reflect temperatures of $4500\text{--}5000^\circ\text{K}$, which are found somewhat higher, near the base of the sun's *chromosphere*.

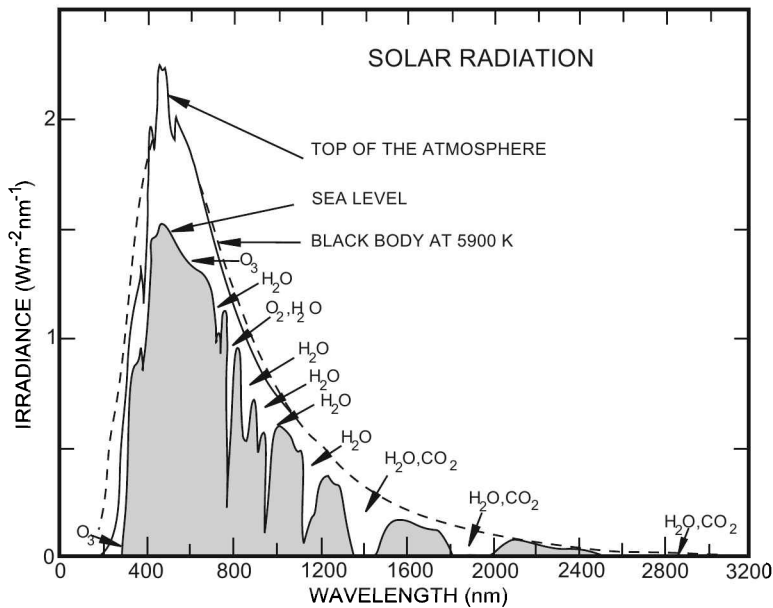


Figure 8.2 Spectrum of SW radiation at the top of the atmosphere (solid) and at the Earth's surface (stippled), compared against the emission spectrum of a blackbody at 6000 K (dashed). Individual absorbing species indicated. Adapted from Coulson (1975).

At the Earth's surface, the wings of the solar spectrum have been mostly clipped off, in the UV by ozone and molecular oxygen and in the IR by absorption bands of water vapor and carbon dioxide. SW radiation is then limited to wavelengths of 0.3–2.4 μm .

Wavelengths shorter than 300 nm account for a small fraction of the SW energy flux. They are absorbed fairly high in the Earth's atmosphere through photodissociation and ionization. As shown in Fig. 8.3, these energetic components penetrate to altitudes in inverse relation to their wavelengths. Wavelengths shorter than 200 nm do not penetrate appreciably below 50 km. Wavelengths shorter than 150 nm are confined even higher, to the thermosphere. Exceptional is the discrete *Lyman- α emission* at 121 nm. It corresponds to transitions from the gravest excited state of hydrogen. The Lyman- α line happens to coincide with an optical window in the banded absorption structure at these wavelengths. This feature enables Lyman- α radiation to penetrate into the lower thermosphere and upper mesosphere, where it is involved in photoionization and dissociation.

The wavelength of SW radiation is also inversely related to the altitude in the solar atmosphere from which it is emitted. As indicated in Fig. 8.4, wavelengths longer than 300 nm originate in the photosphere. Wavelengths shorter than 200 nm are emitted from the chromosphere. Very energetic components, with wavelengths shorter than 50 nm, originate even higher, in the corona. Solar variability increases with altitude in the solar atmosphere. For this reason, temporal variations in the SW spectrum are concentrated in the far UV. Those wavelengths make only a minor contribution to the solar constant.

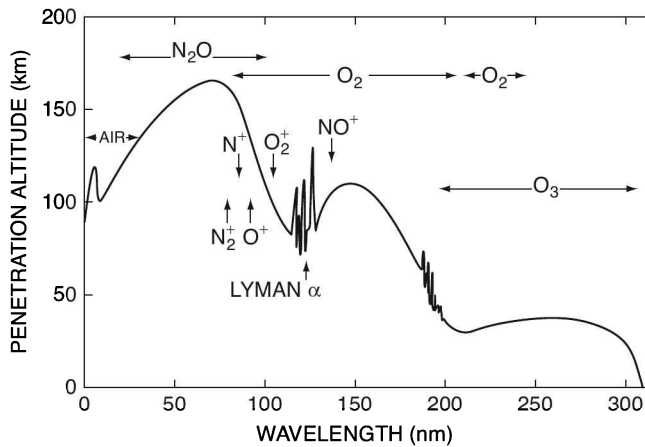


Figure 8.3 Penetration altitude, where SW flux at wavelength λ is attenuated by a factor of e^{-1} . Absorbing species and thresholds for photo-ionization indicated. After Herzberg (1965).

Two phenomena dominate solar variability. The 27-day rotation cycle of the sun leads to variations of several percent in the UV. However, it decreases to less than 1% at wavelengths longer than 250 nm (WMO, 1986). Solar activity also varies with the prevalence of dark regions, *sunspots*. The latter evolve through an 11-year cycle that modulates solar emission. It is noteworthy that, according to proxy evidence, solar activity experienced a distinct minimum during the seventeenth century. The so-called *Maunder minimum* coincided with the Little Ice Age (Sec. 1.6.2). Between extremes of the 11-year solar cycle, *solar-max* and *solar-min*, activity changes noticeably at outer reaches of the sun's atmosphere. Owing to the inverse relationship between solar altitude and the emitted wavelength, SW variability decreases sharply with increasing λ (Fig. 8.4). This sharply limits the altitudes in the Earth's atmosphere that are affected by such variations. At $\lambda = 160$ nm, the peak-to-peak variation in flux between solar-max and solar-min is about 10%. By 300 nm, however, it has decreased to less than 1%. In terms of the solar constant and the majority of energy reaching the Earth's surface, the variation between solar-max and solar-min is of order 0.1% (Willson et al., 1986).

The spectrum of LW radiation emitted to space by the Earth-atmosphere system is presented in Fig. 8.5. Measured by satellite, it too resembles the emission spectrum of a blackbody. However, it corresponds to a temperature of 288 K only at wavelengths in the atmospheric window: 8–12 μm (wavenumbers λ^{-1} of 800–1200 cm^{-1}). Only there does radiation emitted by the Earth's surface pass freely to space. Other wavenumbers are trapped, absorbed, and then re-emitted by overlying water vapor and cloud. Those atmospheric absorbers emit radiation to space at temperatures colder than the Earth's surface.

The strong absorption bands of CO_2 and O_3 sharply reduce emission to space at 15 μm and 9.6 μm , respectively. Outgoing radiation at those wavelengths corresponds to blackbody temperatures that are distinctly colder than that within the atmospheric window, where outgoing radiation emanates directly from the Earth's surface. The 15- μm absorption of CO_2 is positioned near the center of the LW emission spectrum. It takes a bite out of the spectrum that would otherwise be emitted to space.

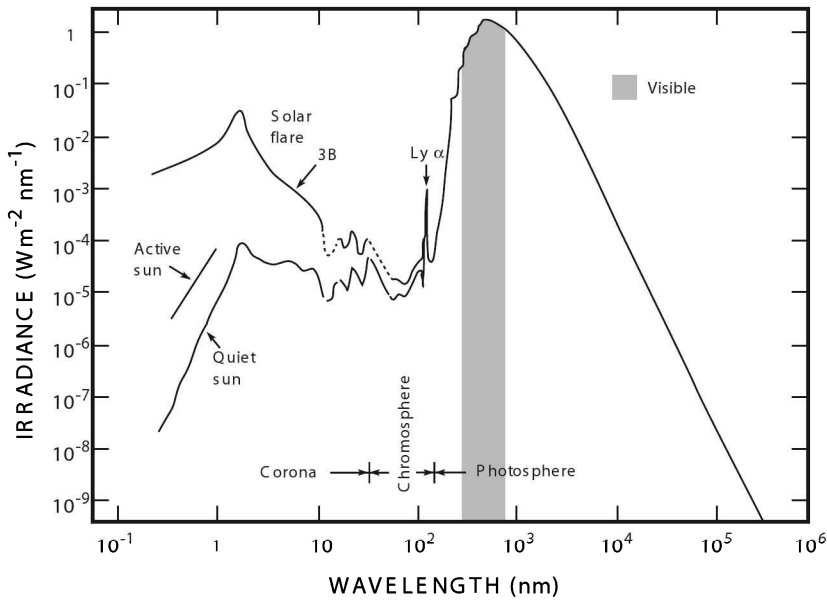


Figure 8.4 Spectrum of solar irradiance (flux) as a function of wavelength under normal and disturbed conditions. Visible band and emission levels in the solar atmosphere indicated. Adapted from Smith and Gottlieb (1974) with permission of Kluwer Academic Publishers.

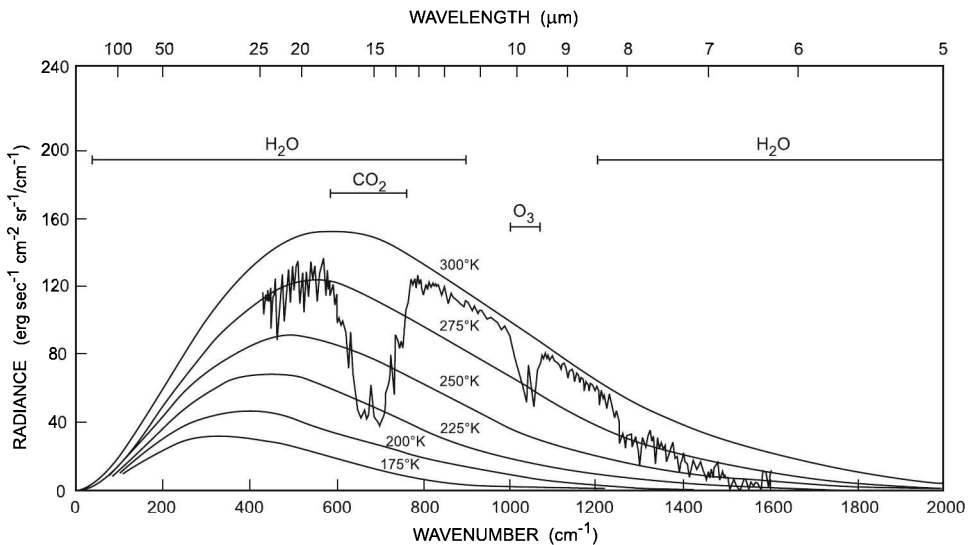


Figure 8.5 Spectrum of outgoing LW radiation over (215 W, 15 N) observed by Nimbus-4 IRIS, as a function of wavenumber λ^{-1} . Blackbody spectra for different temperatures and individual absorbing species indicated. Adapted from Liou (1980).

Radiation at those wavelengths corresponds to a blackbody temperature of about 220 K. That temperature reflects an altitude in the upper troposphere, above most of the column abundance of CO₂ (Fig. 1.7). The 9.6- μm absorption of ozone corresponds to a temperature of about 270 K. That temperature reflects an altitude in the upper stratosphere, above most of the column abundance of ozone. Along with water vapor, these minor constituents block LW radiation that is emitted by the Earth's surface. By absorbing that radiation and re-emitting part of it back downward, they increase surface temperature over what it would be in the absence of an atmosphere (namely, over the effective blackbody temperature of 255 K). Known as the *greenhouse effect*, this process follows from the atmosphere's selective transmissivity: It is transparent to SW radiation that arrives from the sun. However, it is opaque to LW radiation that is re-emitted by the Earth's surface.

8.2 DESCRIPTION OF RADIATIVE TRANSFER

8.2.1 Radiometric quantities

Understanding how radiation influences atmospheric properties requires a quantitative description of radiative transfer. The latter is complicated by the 3-dimensional nature of radiation and its dependence on wavelength and directionality. As illustrated in Fig. 8.6, radiant energy traversing a surface with unit normal \mathbf{n} has contributions from all directions, which are characterized by the unit vector $\hat{\Omega}$. Such multidimensional radiation is termed *diffuse*. If it arrives along a single direction (like incoming solar), it is termed *parallel-beam radiation*. The monochromatic *intensity* or *radiance* is given by $I_\lambda = I_\lambda \hat{\Omega}$. It represents the rate that energy of wavelengths λ to $\lambda + d\lambda$ flows per unit area through an increment of solid angle $d\Omega$ in the direction $\hat{\Omega}$. Having dimensions of $\frac{\text{power}}{\text{wavelength} \cdot \text{area} \cdot \text{steradian}}$, $I_\lambda(\mathbf{x}, \hat{\Omega})$ is a function of position \mathbf{x} and direction $\hat{\Omega}$. It characterizes a pencil of radiation that traverses the surface at a *zenith angle* θ from its normal. Then the rate energy flows in the direction $\hat{\Omega}$, per unit cross-sectional area and wavelength interval, is $I d\Omega$.

The component of intensity normal to the surface is $I_\lambda \cdot \mathbf{n} = I_\lambda(\hat{\Omega} \cdot \mathbf{n}) = I_\lambda \cos\theta$. Integrating it over the half-space of solid angle in the positive \mathbf{n} direction defines the monochromatic *flux* or *irradiance* crossing the surface

$$\begin{aligned} F_\lambda^+ &= \int_{2\pi} I_\lambda \cdot \mathbf{n} d\Omega^+ \\ &= \int_{2\pi} I_\lambda \cos\theta d\Omega^+ \\ &= \int_0^{2\pi} \int_0^{\frac{\pi}{2}} I_\lambda(\phi, \theta) \cos\theta \sin\theta d\theta d\phi. \end{aligned} \quad (8.1)$$

F_λ^+ has dimensions of $\frac{\text{power}}{\text{area} \cdot \text{wavelength}}$. The *total flux* then follows by integrating over wavelength

$$F^+ = \int_0^\infty F_\lambda^+ d\lambda, \quad (8.2)$$

which has the dimensions of energy flux: $\frac{\text{power}}{\text{area}}$. The monochromatic flux F_λ^+ thus represents the spectral density at wavelength λ of the total flux F^+ .

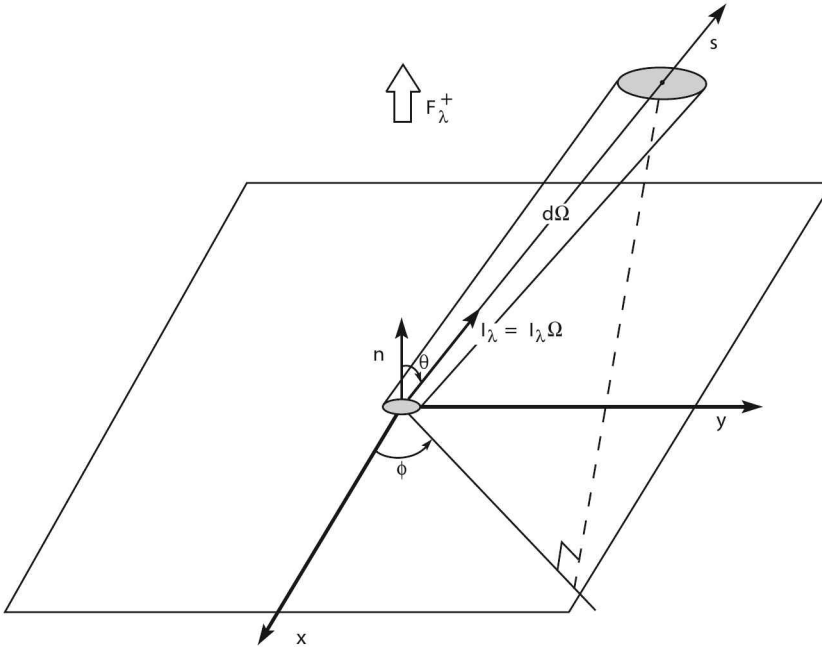


Figure 8.6 A pencil of radiation that occupies the increment of solid angle $d\Omega$ in the direction $\hat{\Omega}$ and traverses a surface with unit normal \mathbf{n} . The monochromatic *intensity* or *radiance* passing through the pencil: $\mathbf{I}_\lambda = I_\lambda \hat{\Omega}$, describes the rate at which energy inside the pencil crosses the surface per unit area, steradian, and wavelength. Integrating the component normal to the surface: $I_\lambda (\hat{\Omega} \cdot \mathbf{n}) = I_\lambda \cos\theta$, over the half-space of 2π steradians in the positive \mathbf{n} direction yields the monochromatic forward *flux* or *irradiance* F_λ^+ .

F_λ^+ describes the forward energy flux in the \mathbf{n} direction. The backward energy flux F_λ^- may be obtained by integrating \mathbf{I} in the negative \mathbf{n} direction over the opposite half-space. Both are nonnegative quantities. The monochromatic *net flux* in the \mathbf{n} direction is then the difference between the forward and backward components

$$F_\lambda = F_\lambda^+ - F_\lambda^- \quad (8.3)$$

Applying (8.1) and (8.3) to the three coordinate directions obtains the net flux vector \mathbf{F}_λ . In the special case of *isotropic* radiation, intensity is independent of direction: $I_\lambda \neq I_\lambda(\hat{\Omega})$. Under these circumstances, (8.1) reduces to

$$F_\lambda^+ = \pi I_\lambda \quad (8.4)$$

Because the forward and backward components are then equal, the net flux in any direction vanishes.

Electromagnetic radiation can interact with matter through three basic mechanisms: absorption, scattering, and emission. A pencil of radiation occupying the solid angle $d\Omega$ is attenuated in proportion to the density and absorption characteristics of the medium through which it passes. Energy can also be scattered out of the solid angle, which likewise attenuates energy flowing through $d\Omega$. Together, absorption and scattering account for the net *extinction* of energy passing through the pencil of

radiation. Conversely, energy emitted into the solid angle or scattered into $d\Omega$ from other directions intensifies the flow of energy through that pencil of radiation. These interactions are governed by a series of laws that describe the transfer of radiation through matter.

8.2.2 Absorption

In the absence of scattering, absorption of energy from a pencil of radiation is expressed by the following principle.

Lambert's Law

The fractional energy absorbed from a pencil of radiation is proportional to the mass traversed by the radiation. For an incremental distance ds , this principle is expressed by

$$\frac{dI_\lambda}{I_\lambda} = -\rho\sigma_{a\lambda} ds, \quad (8.5.1)$$

where the constant of proportionality $\sigma_{a\lambda}$ is the specific *absorption cross section* (also referred to as the *mass absorption coefficient*), which has units of $\frac{\text{area}}{\text{mass}}$. The cross section symbolizes the area of the pencil lost by passing through an increment of mass $\rho dA ds$ or, equivalently, the effective absorbing area of that mass - each for the wavelength λ . The foregoing principle may also be expressed in terms of the particle number density n (e.g., of molecules or aerosol) through which the radiation passes

$$\frac{dI_\lambda}{I_\lambda} = -n\hat{\sigma}_{a\lambda} ds, \quad (8.5.2)$$

where the absorption cross section $\hat{\sigma}_{a\lambda}$ is referenced to an individual particle and has dimensions of area. Thus, $\hat{\sigma}_{a\lambda}$ symbolizes the effective absorbing area posed by a particle in the path of the radiation. In either representation, Lambert's law for attenuation of energy is linear in the intensity I_λ .

The *absorption coefficient*

$$\begin{aligned} \beta_{a\lambda} &= \rho\sigma_{a\lambda} \\ &= n\hat{\sigma}_{a\lambda}, \end{aligned} \quad (8.6.1)$$

has dimensions of inverse length. It measures the characteristic distance over which energy is attenuated. The density in (8.6) corresponds to the mass of absorber. In a mixture, it may follow from several constituent gases. Collecting contributions from different constituents gives

$$\beta_{a\lambda} = \sum_i r_i \rho \sigma_{a\lambda,i} \quad (8.6.2)$$

$$= \sum_i r_i n \hat{\sigma}_{a\lambda,i}, \quad (8.6.3)$$

where r_i denotes the mass mixing ratio of the i th absorbing species.

Integrating Lambert's law along the path of radiation obtains

$$I_\lambda(s) = I_\lambda(0) e^{-\int_0^s \rho \sigma_{a\lambda} ds'}. \quad (8.7)$$

In the absence of scattering and emission, the intensity inside a pencil of radiation decreases exponentially with the *optical path length*

$$\begin{aligned} u(s) &= \int_0^s \rho \sigma_{a\lambda} ds' \\ &= \int_0^s \beta_{a\lambda} ds'. \end{aligned} \quad (8.8)$$

$u(s)$ is the dimensionless distance traversed by radiation, weighted according to the density and absorption cross section of the medium. Because Lambert's law involves no directionality, it applies to flux as well as intensity.

The monochromatic *transmissivity* describes the fraction of incident radiation that remains in the pencil at a given distance. It is given by

$$\begin{aligned} \mathcal{T}_\lambda(s) &= \frac{I_\lambda(s)}{I_\lambda(0)} \\ &= e^{-u(s)}. \end{aligned} \quad (8.9)$$

Transmissivity decreases exponentially with optical path length through an absorbing medium. Conversely, the monochromatic *absorptivity* a_λ represents the fraction of incident radiation that has been absorbed from the pencil during the same traversal. Because

$$\mathcal{T}_\lambda + a_\lambda = 1 \quad (8.10)$$

in the absence of scattering and emission, the absorptivity follows as

$$a_\lambda(s) = 1 - e^{-u(s)}. \quad (8.11)$$

With increasing path length through an absorbing medium, absorptivity increases exponentially, approaching unity. In that limit, the medium is said to be *optically thick*.

8.2.3 Emission

To maintain thermal equilibrium, a substance that absorbs radiant energy must also emit it. Like absorption, the emission of energy into a pencil of radiation is proportional to the mass involved. The basis for describing thermal emission is the theory of blackbody radiation. It was developed by Planck and contemporaries near the turn of the century.

Blackbody radiation is an idealization that corresponds to the energy emitted by an isolated cavity in thermal equilibrium. Such radiation is characterized by the following properties:

- The radiation is uniquely determined by the temperature of the emitter.
- For a given temperature, the radiant energy emitted is the maximum possible at all wavelengths.
- The radiation is isotropic.

In addition to being a perfect emitter, a blackbody is also a perfect absorber: Radiation incident upon a blackbody is completely absorbed at all wavelengths.

Planck's Law

To explain blackbody radiation, Planck postulated that the energy E of molecules is quantized and can undergo only discrete transitions that satisfy

$$\Delta E = \Delta n \cdot h\nu, \quad (8.12)$$

where n is integer, h is Planck's constant (Appendix C), and $\nu = \frac{c}{\lambda}$ is the frequency of electromagnetic radiation emitted or absorbed to accomplish the energy transition ΔE . Hence, the radiation emitted or absorbed by individual molecules is quantized in *photons*, which carry energy in integral multiples of $h\nu$. On this basis, Planck derived the following relationship between the spectrum of intensity B_λ emitted by a population of molecules and their absolute temperature

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5 (e^{\frac{hc}{\lambda T}} - 1)}, \quad (8.13)$$

where K is the Boltzmann constant. Plotted in Fig. 8.7, the *Planck* or *blackbody spectrum* (8.13) increases with temperature. It possesses a single maximum at wavelength λ_m . At wavelengths shorter than λ_m , B_λ decreases sharply. At wavelengths longer than λ_m , it involves a comparatively wide band. Because blackbody radiation is isotropic, the form of the Planck spectrum applies to flux (8.4) as well as to intensity.

Wien's Displacement Law

The wavelength of maximum intensity follows from (8.13) as

$$\lambda_m = \frac{2897}{T} (\mu\text{m}). \quad (8.14)$$

It decreases with increasing temperature of the emitter. This feature of the emission spectrum allows the *brightness temperature* of a body to be inferred from radiation emitted by it. For instance, SW radiation incident on TOA is concentrated in the visible. It peaks near a wavelength of $0.480 \mu\text{m}$ (Fig. 8.2), which corresponds to blue light. Wien's displacement law then implies an effective solar temperature of about 6000°K . On the other hand, the 288°K mean surface temperature of Earth corresponds to maximum emission at a wavelength of about $10 \mu\text{m}$ (Fig. 8.1).

Analogous reasoning can be applied to individual wavelengths via (8.13). For instance, minima in the observed spectrum of outgoing LW radiation (Fig. 8.5) appear at $15 \mu\text{m}$ and $9.6 \mu\text{m}$. They can be identified with emission by CO_2 in the upper troposphere, at the top of its column abundance, and by O_3 in the upper stratosphere, at the top of its column abundance. Each corresponds to a blackbody temperature that is significantly colder than the Earth's surface temperature. In fact, surface temperature is manifest in the spectrum of outgoing LW radiation only in the atmospheric window at $8\text{--}12 \mu\text{m}$.

The Stefan-Boltzmann Law

The total flux emitted by a blackbody follows by integrating over the electromagnetic spectrum. In combination with (8.4), this yields

$$\begin{aligned} F &= \pi B(T) \\ &= \sigma T^4, \end{aligned} \quad (8.15)$$

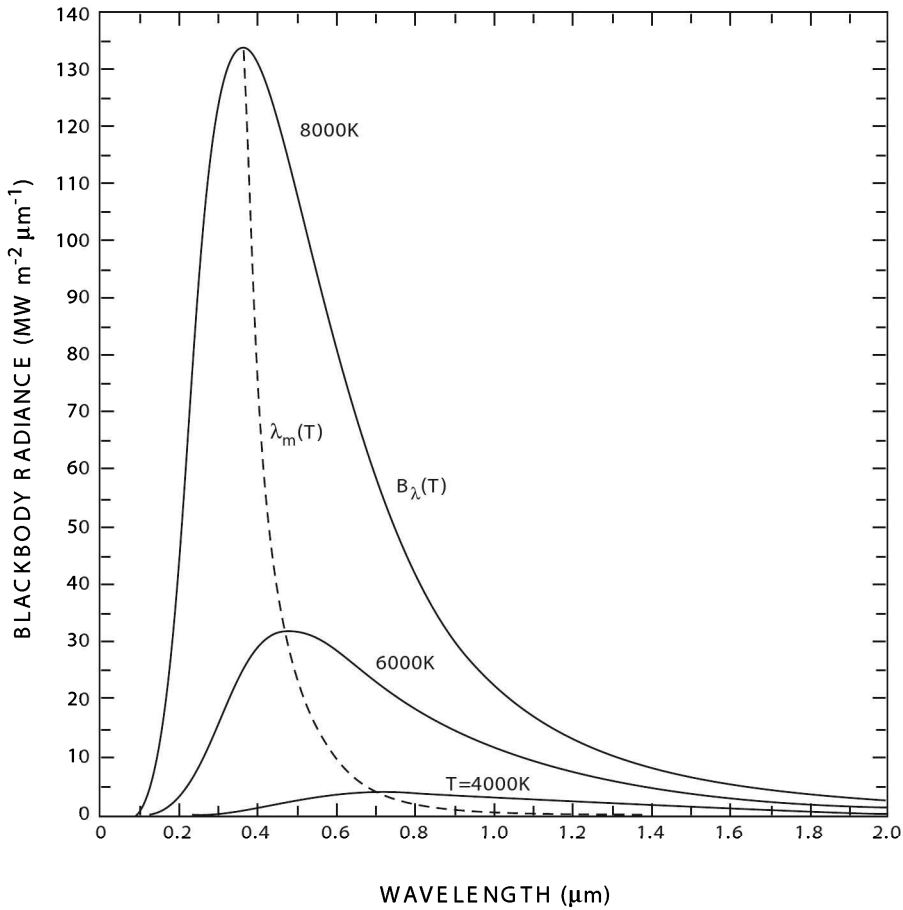


Figure 8.7 Spectra of emitted intensity $B_{\lambda}(T)$ for blackbodies at several temperatures, with wavelength of maximum emission $\lambda_m(T)$ indicated.

where σ is the Stefan-Boltzmann constant (Appendix C).²

The quartic temperature dependence in (8.15) implies LW emission that differs sharply between cold and warm objects. This feature of LW radiation is inherent in IR imagery like that presented in Figs. 1.28 and 1.30. The Sahara desert at 1200 UT is very dark, having temperature in excess of 315°K . It emits LW flux in excess of 550 W m^{-2} . (This strong surface emission is mediated by the atmosphere, which limits the outgoing LW flux at TOA to 350 W m^{-2} or less.) At the same time, deep convective tower over Indonesia has temperature as cold as 180°K . It emits LW flux of only $\sim 60 \text{ W m}^{-2}$. The disparate LW emission by these regions of different temperature symbolizes geographical variations in the Earth's energy budget. It is also symbolic of the diurnal variation

² The Stefan-Boltzmann law and Wien's displacement law predate Planck's postulate of the quantization of energy and his derivation of (8.13). While accounting for these earlier results, Planck's law resolved the failure of classical mechanics to explain the thermal emission spectrum at short wavelengths.

of emission, which follows from changes during the day of surface temperature and cloud cover.

Kirchoff's Law

The preceding laws determine the emission into a pencil of radiation during its passage through a medium that behaves as a blackbody. Real substances absorb and emit radiation at rates smaller than those of a perfect absorber and emitter. A *gray body* absorbs and emits with efficiencies that are independent of wavelength. The *absorptivity* a_λ of a substance is defined as the ratio of the intensity it absorbs to $B_\lambda(T)$. According to (8.11), the absorptivity of a layer approaches unity with increasing optical path length. At that point, it behaves as a blackbody. Similarly, the *emissivity* ϵ_λ is defined as the ratio of intensity emitted by a substance to $B_\lambda(T)$. According to the properties of blackbody radiation, it too must approach unity with increasing optical path length.

Consider two infinite plates that communicate with one another only radiatively: one a blackbody and the other a gray body with constant absorptivity a and emissivity ϵ . Suppose the plates are in thermal equilibrium, so they absorb and emit energy in equal proportion. Suppose further that the plates have different temperatures. Introducing a conducting medium between the plates would then drive the system out of thermal equilibrium by permitting heat to flow from the warmer to the colder plate. To restore thermal equilibrium, radiation would then have to transfer energy from the colder plate back to the warmer plate. This would violate the second law of thermodynamics. It follows that the *radiative equilibrium temperature* of the gray plate, that temperature at which it emits energy at the same rate as it absorbs energy from its surroundings, must be identical to that of the black plate.

For the gray plate to be in equilibrium, the flux of energy it emits: $\epsilon\sigma T^4$, must equal the flux of energy it absorbs from the black plate: $a\sigma T^4$. Applying this reasoning to substances of arbitrary monochromatic absorptivity and emissivity leads to the conclusion

$$\epsilon_\lambda = a_\lambda. \quad (8.16)$$

Known as *Kirchoff's law*, (8.16) asserts that a substance emits radiation at each wavelength as efficiently as it absorbs it.

In general, the radiative efficiency of a substance varies with wavelength. As illustrated in Prob. 8.10, an illuminated surface with absorptivity a_{SW} in the SW and a_{LW} in the LW has a radiative-equilibrium temperature that varies according to the ratio $\frac{a_{SW}}{a_{LW}}$. A surface such as snow absorbs weakly in the visible but strongly in the IR. It thus has a lower radiative-equilibrium temperature than a gray surface, for which the absorptivity is independent of wavelength. In a similar manner, the selective transmissivity of the atmosphere increases the radiative-equilibrium temperature of the Earth's surface over what it would be in the absence of an atmosphere. To SW radiation, the atmosphere is transparent. Passing freely through the atmosphere, that radiation is absorbed at the Earth's surface. To LW radiation, the atmosphere is opaque. Re-emitted by the Earth's surface to preserve thermal equilibrium, that radiation is trapped by the overlying atmosphere, which absorbs at those wavelengths (Prob. 8.21). To offset reduced transmission, the surface must emit more LW radiation, which requires a

higher temperature (8.15). Surface temperature must therefore increase, just enough for the LW radiation that is rejected to space at TOA to balance the SW radiation that is absorbed. The increased surface temperature represents the greenhouse effect. It is related directly to the difference of atmospheric opacity to SW and LW radiation (Prob. 8.22).

Kirchoff's law is predicated on a state of *local thermodynamic equilibrium* (LTE), wherein temperature is uniform and radiation is isotropic within some small volume. Those conditions are satisfied when energy transitions are dominated by molecular collision. For the most important radiatively active gases, they hold at pressures greater than 0.01 hPa (e.g., at altitudes below 60 km). At greater height, LTE breaks down because the interval between collisions is no longer short compared with the lifetime of excited states associated with absorption and emission. Kirchoff's law is then invalid.

With (8.16), the fractional energy emitted into a pencil of radiation along an incremental distance ds may be expressed

$$\begin{aligned} \frac{dI_\lambda}{B_\lambda} &= d\epsilon_\lambda = da_\lambda \\ &= \rho\sigma_{a\lambda} ds \end{aligned}$$

by Lambert's law. Then

$$dI_\lambda = \rho\sigma_{a\lambda} J_\lambda ds, \quad (8.17)$$

where

$$J_\lambda = B_\lambda(T) \quad (8.18)$$

defines the monochromatic *source function* in the absence of scattering. Equation (8.18) holds under the conditions of Kirchoff's law, namely, under LTE. Otherwise, the contribution to the source function from emission is more complex.

8.2.4 Scattering

Beyond absorption and emission, energy passing through a pencil of radiation is also modified by scattering. The scattering process refers to the extraction and subsequent re-emission of energy by matter. A population of molecules possesses, in addition to its translational energy associated with temperature, electronic, vibrational, and rotational energies. Internal to individual molecules, those forms of energy can be excited by absorbing a photon. The absorbed energy can then be released in several ways. The simplest is when the excited internal energy is converted into translational energy through molecular collision, or "thermalized." This corresponds to thermal absorption of radiation. Thermal emission occurs through the reverse process. The excited energy can also be re-emitted by molecules - in wavelengths and directions different from those of the incident radiation (Fig. 8.8). This constitutes scattering. Analogous interactions occur between radiation and atmospheric aerosol.

The foregoing process is described in terms of the monochromatic *scattering cross section* $\sigma_{s\lambda}$, which symbolizes the fractional area removed from a pencil of radiation

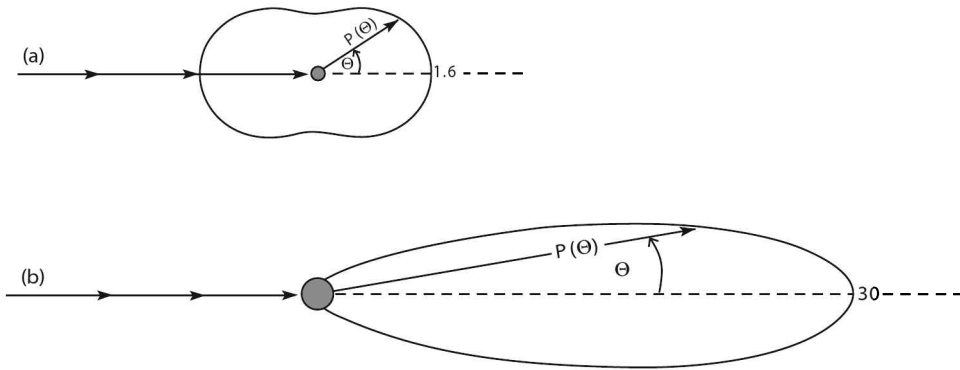


Figure 8.8 Angular distribution of radiation scattered from (a) small particles (of radius $a \ll \lambda$), which is representative of *Rayleigh scattering* of SW radiation by air molecules, and (b) large particles ($a \gg \lambda$), which is representative of *Mie scattering* of SW radiation by cloud droplets (Sec. 9.4.1). Phase function P is plotted in terms of the scattering angle Θ and in (b) for a scattering population with the refractive index of water and an effective size parameter $x_g = 2\pi \frac{a_g}{\lambda} = 5$. Note: The compressed scale in (b) implies that energy redirected by large particles is dominated by forward scattering. Larger particles produce even stronger forward scattering (cf. Fig. 9.30).

through scattering. As absorption and scattering are both linear in I_λ , their effects are additive. The monochromatic *extinction cross section* is defined by

$$k_\lambda = \sigma_{a\lambda} + \sigma_{s\lambda}. \quad (8.19.1)$$

The *extinction coefficient* then follows as

$$\beta_{e\lambda} = \rho k_\lambda, \quad (8.19.2)$$

where ρ is understood to refer to the optically active species. Lambert's law (8.5) and its consequences for attenuation then hold with k_λ in place of $\sigma_{a\lambda}$.

Scattering also modifies how energy is introduced into a pencil of radiation. This makes the source function more complex than the contribution from emission alone (8.18). While scattering removes radiation from one direction, it introduces it into other directions. Alternatively, photons can be scattered into a particular solid angle $d\Omega$ from all directions. If photons are introduced into the pencil of radiation through only one encounter with a particle, the process is termed *single scattering*. If more encounters are involved, the process is termed *multiple scattering*.

The *single scattering albedo*

$$\omega_\lambda = \frac{\sigma_{s\lambda}}{k_\lambda} \quad (8.20)$$

represents the fraction of radiation lost through extinction that is scattered out of a pencil of radiation. Then

$$1 - \omega_\lambda = \frac{\sigma_{a\lambda}}{k_\lambda} \quad (8.21)$$

represents the fraction lost through extinction that is absorbed from the pencil of radiation. The directionality of the scattered component is described in the *phase function* $P_\lambda(\hat{\Omega}, \hat{\Omega}')$. It represents the fraction of radiation scattered by an individual particle from the direction $\hat{\Omega}'$ into the direction $\hat{\Omega}$. If the phase function is normalized according to

$$\frac{1}{4\pi} \int_{4\pi} P_\lambda(\hat{\Omega}, \hat{\Omega}') d\Omega' = 1, \quad (8.22)$$

$\omega_\lambda \frac{P_\lambda(\hat{\Omega}, \hat{\Omega}')}{4\pi} d\Omega'$ represents the fraction of radiation lost through extinction from the pencil in the direction $\hat{\Omega}'$ that is scattered into another pencil in the direction $\hat{\Omega}$. The scattered contribution to the source function is then

$$J_{s\lambda} = \frac{\omega_\lambda}{4\pi} \int_{4\pi} I_\lambda(\hat{\Omega}') P_\lambda(\hat{\Omega}, \hat{\Omega}') d\Omega'. \quad (8.23)$$

Combining contributions from emission and scattering yields the total source function

$$J_\lambda = (1 - \omega_\lambda) B_\lambda(T) + \frac{\omega_\lambda}{4\pi} \int_{4\pi} I_\lambda(\hat{\Omega}') P_\lambda(\hat{\Omega}, \hat{\Omega}') d\Omega'. \quad (8.24)$$

8.2.5 The Equation of Radiative Transfer

Collecting the extinction of energy (8.19) and the introduction of energy (8.17) yields the net rate at which the intensity inside a pencil of radiation changes in the s direction

$$\frac{dI_\lambda}{\rho k_\lambda ds} = -I_\lambda + J_\lambda, \quad (8.25)$$

which is the *radiative transfer equation* in general form. In the absence of scattering, (8.25) reduces to

$$\frac{dI_\lambda}{\rho k_\lambda ds} = -I_\lambda + B_\lambda(T), \quad (8.26)$$

where k_λ then equals the absorption cross section $\sigma_{a\lambda}$. The *optical thickness* is defined by

$$\begin{aligned} \chi_\lambda(s) &= \int_s^0 \rho k_\lambda ds' \\ &= -u(s). \end{aligned} \quad (8.27.1)$$

It increases in the negative s direction (i.e., opposite to $\hat{\Omega}$). Then

$$d\chi_\lambda = -\rho k_\lambda ds, \quad (8.27.2)$$

casts the energy budget for a pencil of radiation into the canonical form

$$\frac{dI_\lambda}{d\chi_\lambda} = I_\lambda - B_\lambda(T), \quad (8.28)$$

which is known as *Schwartzchild's equation*.

The solution of Schwartzchild's equation may be expressed formally with the aid of an integrating factor. Multiplying by $e^{-\chi_\lambda}$ allows (8.28) to be written

$$\frac{d}{d\chi_\lambda} [e^{-\chi_\lambda} I_\lambda] = -e^{-\chi_\lambda} B_\lambda(T).$$

Upon integrating from 0 to $\chi_\lambda(s)$, it yields

$$I_\lambda(s) = I_\lambda(0)e^{\chi_\lambda(s)} - \int_0^{\chi_\lambda(s)} B_\lambda[T(\chi'_\lambda)]e^{\chi_\lambda(s)-\chi'_\lambda} d\chi'_\lambda. \quad (8.29)$$

The first term in (8.29) describes an exponential decrease with optical path length of the incident intensity $I_\lambda(0)$, as embodied in Lambert's law (8.7). The second term describes the cumulative emission and absorption between 0 and $\chi_\lambda(s)$. If the properties T , ρ , and k_λ are known as functions of s , (8.29) uniquely determines the intensity along the path of radiation.

8.3 ABSORPTION CHARACTERISTICS OF GASES

8.3.1 Interaction between radiation and molecules

The absorptivity of a medium approaches unity with increasing optical path length, irrespective of wavelength (Sec. 8.2.1). Along finite path lengths (e.g., through an atmosphere of bounded mass), absorption may be large at some wavelengths and small at others, according to the optical properties of the medium. For a gas, the absorption spectrum a_λ is concentrated in a complex array of lines that correspond to transitions between the discrete electronic, vibrational, and rotational energy levels of molecules. Electronic transitions are stimulated by radiation at UV and visible wavelengths. Vibrational and rotational transitions occur at IR wavelengths. At pressures greater than about 0.1 hPa (e.g., below 60 km), the internal energy acquired by absorbing a photon is quickly thermalized. In addition to discrete absorption characteristics, a continuum of absorption occurs at shorter wavelengths in connection with photodissociation and ionization of molecules. The latter occur at wavelengths of X-ray, UV, and, to a lesser degree, of visible radiation. They are possible for all λ shorter than the threshold to break molecular and electronic bonds.

Figure 8.9 displays absorption spectra in the IR for optically active gases corresponding to a vertical path through the atmosphere. All of the species are minor constituents of air. The most important are water vapor, carbon dioxide, and ozone. Each of these triatomic molecules is capable of undergoing simultaneous vibration-rotation transitions. They produce a clustering of absorption lines. At coarse resolution, like spectra in Figs. 8.1 and 8.9, those clusters appear as continuous bands of absorption. Water vapor absorbs strongly in a broad band centered near $6.3 \mu\text{m}$ and in another at $2.7 \mu\text{m}$. Both bands correspond to transitions from vibrationally excited states: $6.3 \mu\text{m}$ absorption to the ν_2 vibrational mode (Fig. 8.10a), whereas the band at $2.7 \mu\text{m}$ involves both ν_1 and ν_3 modes of vibration. Rotational transitions of H_2O are less energetic. They lead to absorption at wavelengths longer than $12 \mu\text{m}$ (cf. Fig. 8.1). Carbon dioxide is excited vibrationally by wavelengths near $15 \mu\text{m}$ and also near $4.3 \mu\text{m}$. The former corresponds to the transverse mode ν_2 (Fig. 8.10b), whereas the latter corresponds to the longitudinal vibration ν_3 . Ozone absorbs strongly at wavelengths near $9.6 \mu\text{m}$ in connection with vibrational transitions. This absorption band coincides with

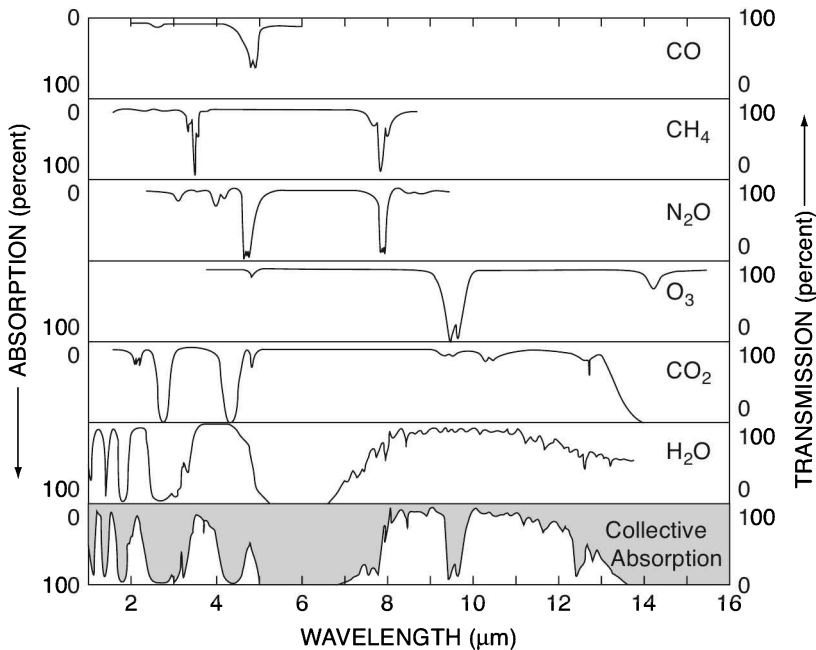
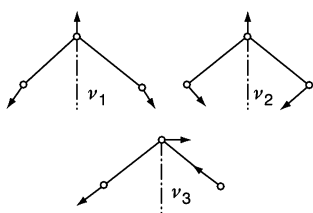


Figure 8.9 Absorption spectra for LW radiation passing vertically through the atmosphere as contributed by strong absorbing gases. Adapted from *Handbook of Geophysics and Space Environment* (1965).

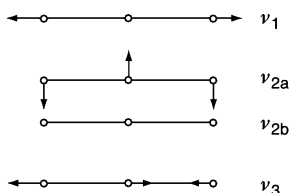
the atmospheric window at 8–12 μm (Fig. 8.1). It therefore enables stratospheric ozone to interact radiatively with the troposphere and the Earth's surface. Except for ozone, most of the absorption by these species takes place in the troposphere, where absolute concentrations are large (cf. Figs. 8.1b,c). Nitrous oxide, methane, carbon monoxide, and CFC-11 and -12 also have absorption lines within the range of wavelength shown in Fig. 8.9. However, they are of secondary importance.

At higher resolution, IR absorption features in Fig. 8.9 actually comprise a complex array of lines. Figure 8.11 illustrates this in high-resolution spectra for the 15 μm band of CO_2 (Fig. 8.11a) and for the rotational band of H_2O at wavelengths of 27–31 μm (Fig. 8.11b). Absorption lines in the 15- μm band of CO_2 are spaced regularly in wavenumber. By contrast, the rotational band of water vapor involves a random distribution of absorption lines. In principle, all such lines must be accounted for in radiative calculations. However, their number and complexity make *line-by-line calculations* impractical for most applications. Instead, radiative calculations rely on *band models*. The latter represent the gross characteristics of absorption spectra in particular ranges of wavelength. Under cloud-free conditions, they provide an economical yet accurate alternative (e.g., Goody et al., 1989; Mlawer et al., 1997).

Similar features characterize the absorption spectrum at visible and UV wavelengths, where radiation interacts with molecular oxygen and ozone. In addition to discrete features, the more energetic radiation at these wavelengths produces continuous bands of absorption in connection with photodissociation and ionization of molecules. Figure 8.12 shows the absorption cross section $\sigma_{a\lambda}$ for O_2 in the UV. The *Herzberg continuum* lies at wavelengths of 242–200 nm, where O_2 is dissociated into



(a)



(b)

Figure 8.10 Normal modes of vibration for triatomic molecules corresponding to (a) H₂O and O₃, in which atoms are configured triangularly, and (b) CO₂, in which atoms are configured longitudinally. After Herzberg (1945).

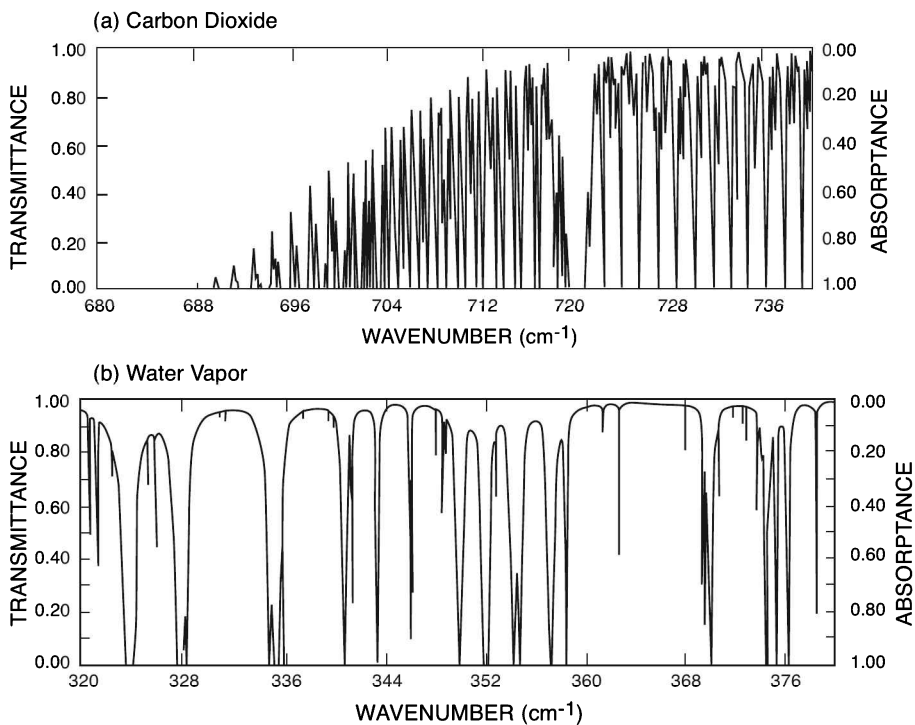


Figure 8.11 Absorption spectra in (a) the 15- μ m band of CO₂ and (b) the rotational band of H₂O at 27-31 μ m. Adapted from McClatchey and Selby (1972).

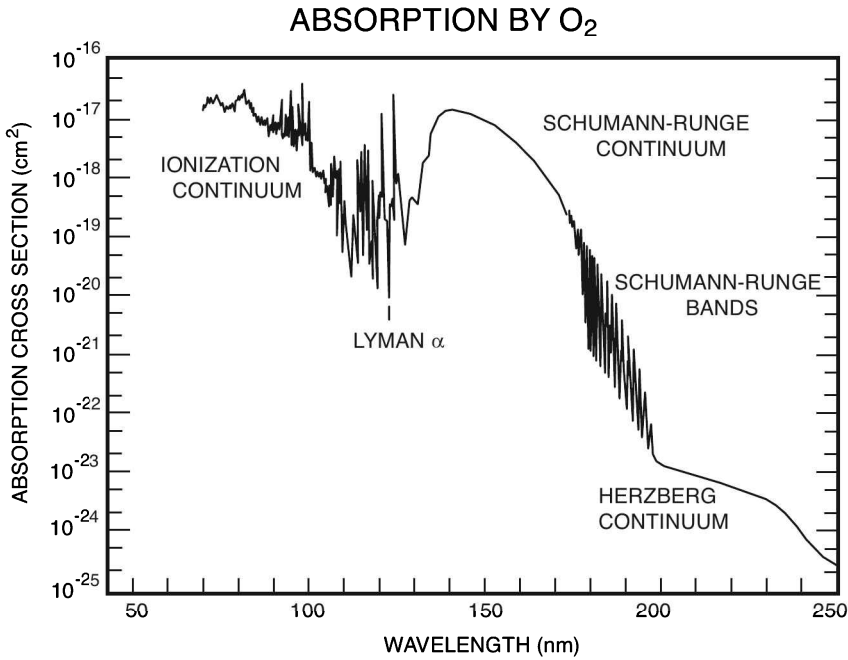


Figure 8.12 Absorption cross section as a function of wavelength for molecular oxygen. After Brasseur and Solomon (1986). Reprinted by permission of Kluwer Academic Publishers.

two ground-state oxygen atoms. At shorter wavelengths, the absorption spectrum is marked by the discrete *Schumann-Runge bands*, where O₂ is vibrationally excited. These excited bound states are unstable; they eventually produce two ground-state oxygen atoms.

Each of the vibrational bands in the Schumann-Runge system is actually comprised of many rotational lines, which are shown at high resolution in Fig. 8.13. The rotational line structure is fairly regular at long wavelength. However, it gradually degenerates to an almost random distribution at short wavelength. Wavelengths shorter than 175 nm are absorbed in the *Schumann-Runge continuum* (Fig. 8.12), in which O₂ is dissociated into two oxygen atoms (one electronically excited). At still shorter wavelengths, absorption exhibits an irregular banded structure that eventually merges into the ionization continuum at wavelengths shorter than 102 nm.

Actual absorption by individual bands in Fig. 8.12 varies strongly with altitude. Even though it has the smallest values of $\hat{\sigma}_{a\lambda}$, the Herzberg continuum at 200–242 nm dominates absorption by O₂ up to 60 km – because shorter wavelengths have already been removed at higher altitude (Fig. 8.3). Above 60 km, the Schumann-Runge bands are dominant. The Schumann-Runge continuum prevails in the thermosphere, where very energetic UV is present.

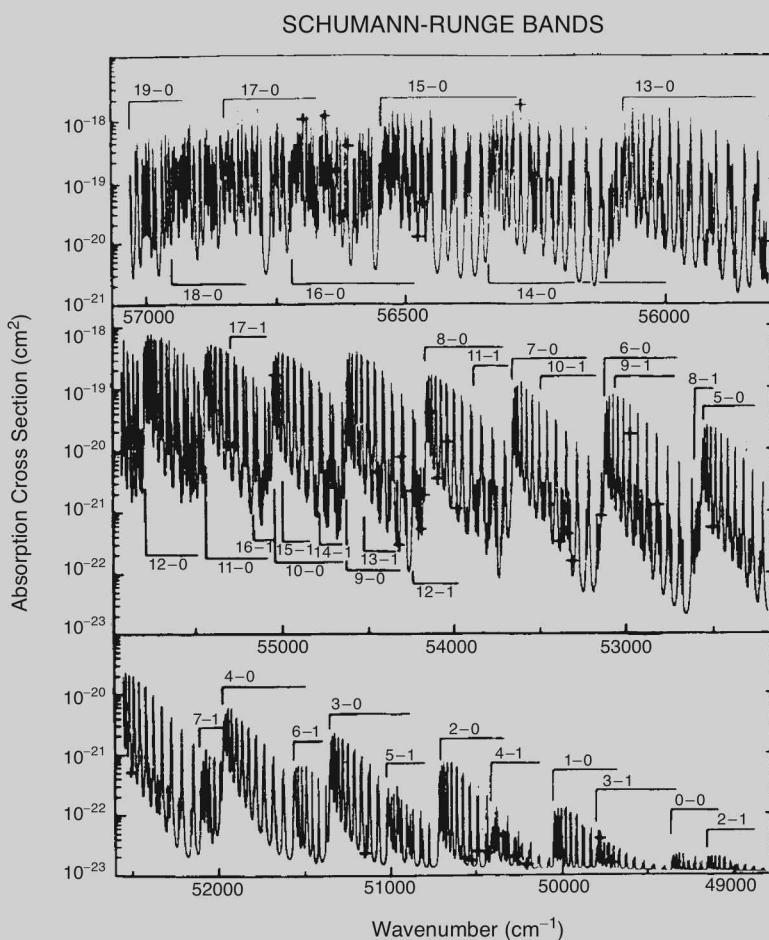


Figure 8.13 Absorption cross section in the Schumann-Runge absorption bands of molecular oxygen. Shown at high spectral resolution. After Kocharts (1971). Reprinted by permission of Kluwer Academic Publishers.

Ozone absorbs at longer wavelengths, which penetrate to lower altitudes. The primary absorption of UV by ozone is in the *Hartley band* at wavelengths of 200–310 nm (Fig. 8.14). At long wavelength, the Hartley band merges with the *Huggins bands* at 310–400 nm. Ozone also absorbs in the visible in the *Chappuis band* at wavelengths of 400–850 nm, which is important at altitudes below 25 km. In all of these bands, absorption follows from photodissociation of ozone. The Hartley and Chappuis bands are continuous. Conversely, the Huggins bands involve a spectrum of diffuse lines. The Chappuis band is weaker than the Hartley and Huggins bands. However, its importance is underscored by its overlap with the peak of the SW spectrum. For this reason, O_3 is rapidly photo-dissociated in sunlight – even at the Earth’s surface. This property couples ozone in the stratosphere to snow and cloud cover in the troposphere, which magnify the reflected SW flux.

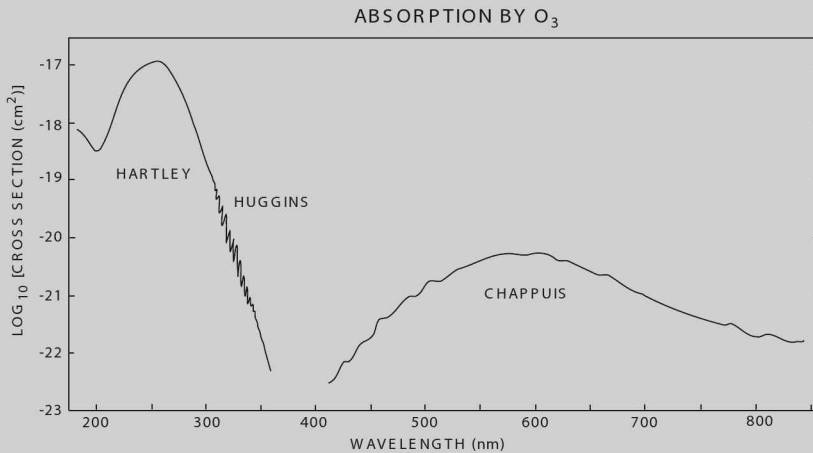


Figure 8.14 Absorption cross section of ozone, as a function of wavelength. *Sources:* Andrews et al. (1987) and WMO (1986).

Absorption and therefore radiative heating follow from the cross sections of individual species, weighted by their respective mixing ratios (Fig. 8.15). Absorption at wavelengths longer than 200 nm is dominated by photodissociation of O₃. Absorption at shorter wavelengths is dominated by photodissociation of O₂. Owing to the inverse relationship between wavelength and penetration altitude (Fig. 8.3), these bands influence different levels. Absorption by O₃ in the Hartley and Huggins bands at $\lambda > 200$ nm dominates in the stratosphere and mesosphere. There, it provides the primary source of heating. Ozone absorption in the Chappuis band at $\lambda > 400$ nm becomes important below 25 km. On the other hand, absorption by O₂ at $\lambda < 200$ nm prevails only only above 60 km. Despite its secondary contribution to absorption, photodissociation of O₂ plays a key role in the energetics of the stratosphere and mesosphere because it produces atomic oxygen, which supports ozone formation (1.27).

8.3.2 Line broadening

Absorption lines in Figs. 8.11 and 8.13 are not truly discrete. Rather, each occupies a finite band of wavelength, due to practical considerations surrounding molecular absorption and emission. The spectral width of an absorbing line is described in terms of a shape factor. The absorption cross section at frequency ν is expressed

$$\sigma_{av} = S f(\nu - \nu_0), \quad (8.30.1)$$

where

$$S = \int \sigma_\nu d\nu \quad (8.30.2)$$

is the *line strength*, ν_0 is the line center, and the shape factor f accounts for line broadening.

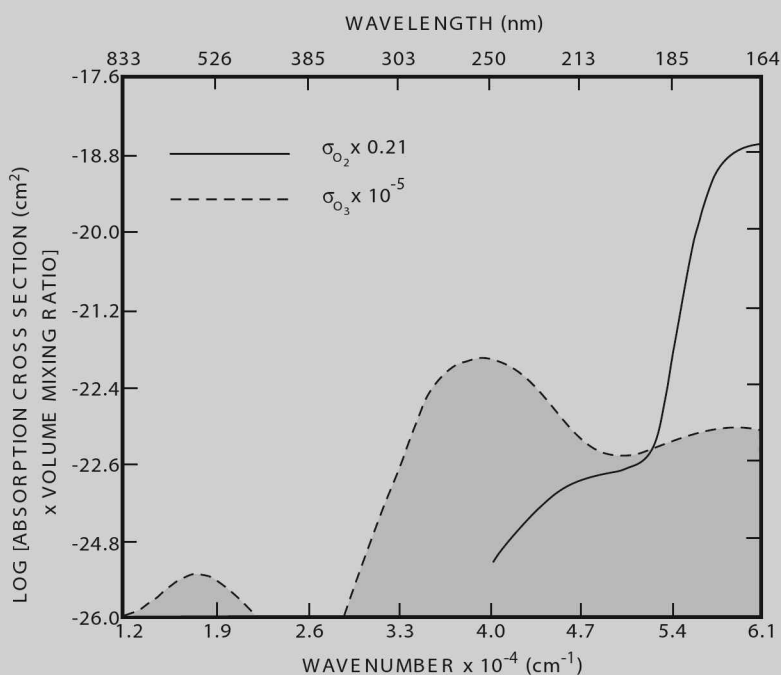


Figure 8.15 Relative contributions to absorption cross section from ozone and molecular oxygen. Adapted from Brasseur and Solomon (1984). Reprinted by permission of Kluwer Academic Publishers.

A fundamental source of line width is *natural broadening*, which results from the finite lifetime of excited states. Perturbations to the radiation stream then introduce a natural width that has the *Lorentz line shape*

$$f_L(\nu - \nu_0) = \frac{\alpha_L}{\pi[(\nu - \nu_0)^2 + \alpha_L^2]}, \quad (8.31.1)$$

where

$$\alpha_L = (2\pi\bar{t})^{-1} \quad (8.31.2)$$

is the half-width at half-power of the line (Fig. 8.16) and \bar{t} is the mean lifetime of the excited state. For vibrational and rotational transitions in the IR, natural broadening is insignificant compared with other sources of line width. The two most important stem from the random motion of molecules and collisions between them.

Collisional or pressure broadening results from perturbations to the absorbing or emitting molecules. Introduced through encounters with other molecules, those perturbations destroy the phase coherence of radiation. The spectral width introduced by collision is modeled in the Lorentz line shape (8.31.1), but with the collisional half-width

$$\alpha_c = \alpha_0 \left(\frac{p}{p_0} \right) \left(\frac{T_0}{T} \right)^{\frac{1}{2}}, \quad (8.32)$$

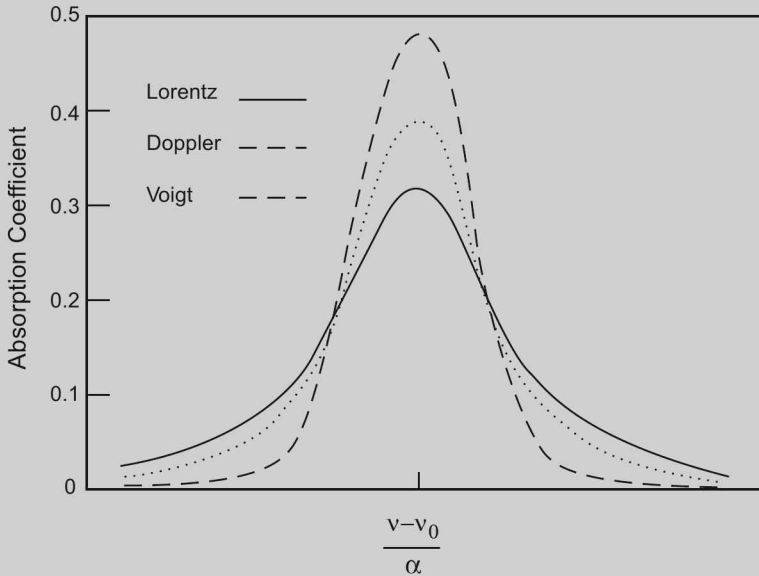


Figure 8.16 Lorentz, Doppler, and Voigt shape factors describing finite spectral line width, plotted as functions of $\frac{v-v_0}{\alpha}$, with α the half width of each line shape. Curves normalized to achieve the same area under each. Reconstructed from Andrews et al. (1987).

where $\alpha_0 \cong 0.1 \text{ cm}^{-1}$ is the half-width at standard temperature and pressure T_0 , p_0 . From kinetic theory, α_c is inversely proportional to the mean time between collisions.

Molecular motion v along the line of sight introduces another source of line width. *Doppler broadening* follows from the frequency shift

$$v = v_0 \left(1 \pm \frac{v}{c} \right) \quad (8.33.1)$$

of individual molecules undergoing emission or absorption. In concert with the Boltzmann probability distribution

$$P(v) = \sqrt{\frac{m}{2\pi KT}} e^{-\frac{mv^2}{2KT}}, \quad (8.33.2)$$

where m is the molecular mass, (8.33.1) leads to the shape factor

$$f_D(v - v_0) = \frac{1}{\alpha_D \sqrt{\pi}} e^{-\left(\frac{v-v_0}{\alpha_D}\right)^2}, \quad (8.34.2)$$

where

$$\alpha_D = \frac{v_0}{c} \sqrt{\frac{2KT}{m}} \quad (8.34.2)$$

is the Doppler half-width divided by a factor of $\sqrt{\ln 2}$. The Doppler line shape is illustrated in Fig. 8.16, along with the collisional line shape for $\alpha_D = \alpha_c$. Unlike the collisional half-width, α_D does not depend on pressure.

Below 30 km, the pressure dependence in α_c makes collisional broadening dominant for IR bands of CO₂ and H₂O. At greater height, Doppler and natural broadening become significant at wavelengths in the visible and UV. Natural broadening of the Schumann-Runge bands of O₂ becomes comparable to collisional broadening in the upper stratosphere and mesosphere. Those two sources of line width are treated jointly in the *Voigt line shape*, which is superposed in Fig. 8.16. When plotted as functions of $\frac{\nu-\nu_0}{\alpha}$, with α the half width of each line shape, the Doppler profile is dominant near the line center, whereas the Lorentz profile prevails in the wings of the line. The Voigt profile then gives a line shape between the two.

8.4 RADIATIVE TRANSFER IN A PLANE PARALLEL ATMOSPHERE

In a stratified atmosphere, absorbers vary sharply with height. It is therefore convenient to treat radiative transfer within the framework of a *plane parallel atmosphere*, wherein

- Curvature associated with sphericity of the Earth is ignored.
- The medium is regarded as horizontally homogeneous and the radiation field horizontally isotropic.

Then, along a slant path ds (Fig. 8.17), a pencil of radiation inclined from the vertical at a zenith angle θ traverses an atmospheric layer of thickness

$$dz = \mu ds, \quad (8.35.1)$$

where

$$\mu = \cos\theta. \quad (8.35.2)$$

It is convenient to introduce the *optical depth*

$$\begin{aligned} \tau_\lambda &= \int_z^\infty \rho k_\lambda dz' \\ &= \frac{1}{g} \int_0^p k_\lambda dp' = \mu \chi_\lambda, \end{aligned} \quad (8.36)$$

which is measured downward from the top of the atmosphere. The energy budget for a pencil of radiation (8.28) is then transformed into

$$\mu \frac{dI_\lambda}{d\tau_\lambda} = I_\lambda - J_\lambda, \quad (8.37)$$

where $I_\lambda = I_\lambda(\tau_\lambda, \mu)$ and $J_\lambda = J_\lambda(\tau_\lambda, \mu)$. Proceeding as in the treatment of (8.28) leads to a formal expression for upwelling radiation ($0 < \mu \leq 1$) in terms of that at the surface ($\tau_\lambda = \tau_{s\lambda}$)

$$I_\lambda(z, \mu) = I_\lambda(\tau_{s\lambda}, \mu) e^{\frac{\tau_\lambda(z) - \tau_{s\lambda}}{\mu}} + \int_{\tau_\lambda(z)}^{\tau_{s\lambda}} J_\lambda(\tau'_\lambda, \mu) e^{\frac{\tau_\lambda(z) - \tau'_\lambda}{\mu}} \frac{d\tau'_\lambda}{\mu} \quad 0 < \mu \leq 1 \quad (8.38.1)$$

and for downwelling radiation ($-1 \leq \mu < 0$) in terms of that at TOA ($\tau_\lambda = 0$)

$$I_\lambda(z, \mu) = I_\lambda(0, \mu) e^{\frac{\tau_\lambda(z)}{\mu}} - \int_0^{\tau_\lambda(z)} J_\lambda(\tau'_\lambda, \mu) e^{\frac{\tau_\lambda(z) - \tau'_\lambda}{\mu}} \frac{d\tau'_\lambda}{\mu} \quad -1 \leq \mu < 0. \quad (8.38.2)$$

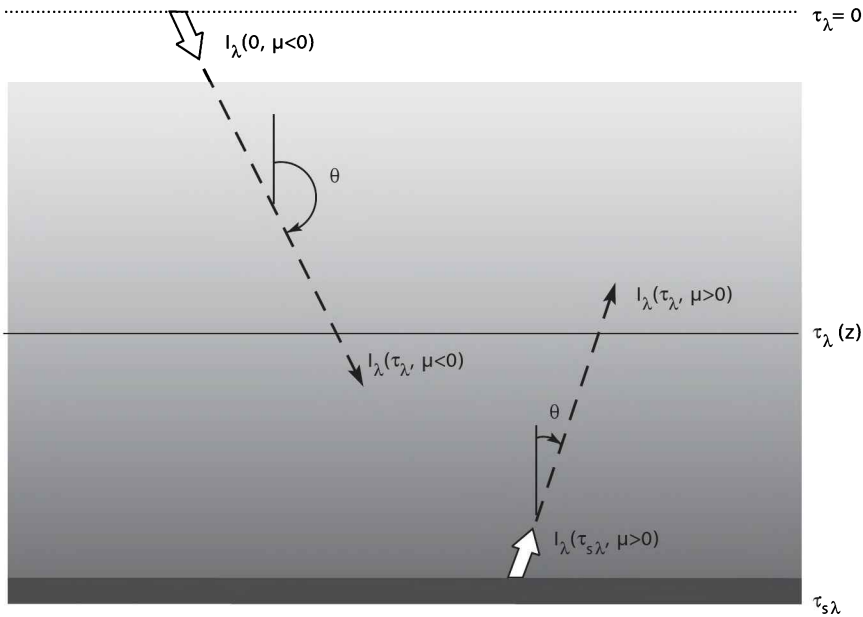


Figure 8.17 Plane parallel atmosphere, in which a pencil of radiation is inclined at the zenith angle $\theta = \cos^{-1}\mu$. Elevation is measured by the optical depth for a given wavelength: τ_λ , which increases downward from zero at the top of the atmosphere to a surface value of $\tau_{s\lambda}$.

As before, the first terms on the right hand sides of (8.38) describe the extinction of radiation emanating from the boundaries. The second terms account for the cumulative emission, absorption, and scattering into the pencil between the boundaries and the height z . If I_λ is known at the upper and lower boundaries and if J_λ can be specified in terms of known properties, (8.38) determines the radiation field throughout the atmosphere. Alternatively, if J_λ , ρ , and k can all be specified in terms of temperature (e.g., in the absence of scattering (8.18) and under hydrostatic equilibrium (1.17)), (8.38) determines the thermal structure of the atmosphere. The expressions for the radiation field then represent integral equations for the temperature distribution $T(z)$. Their solution defines the *radiative-equilibrium thermal structure*, that for which radiative components of the energy budget are in balance.

The upwelling and downwelling fluxes are give by

$$\begin{aligned}
 F_\lambda^\uparrow(\tau_\lambda) &= 2\pi \int_0^1 I_\lambda(\tau_\lambda, \mu) \mu d\mu \\
 F_\lambda^\downarrow(\tau_\lambda) &= -2\pi \int_{-1}^0 I_\lambda(\tau_\lambda, \mu) \mu d\mu.
 \end{aligned}
 \tag{8.39}$$

Horizontal homogeneity and isotropy make these the essential descriptors of radiative transfer in a plane parallel atmosphere. Integrating (8.38) over zenith angle and

incorporating (8.4) obtains

$$F_{\lambda}^{\uparrow}(\tau_{\lambda}) = 2\pi \int_0^1 I_{\lambda}(\tau_{s\lambda}, \mu) e^{\frac{\tau_{\lambda}-\tau_{s\lambda}}{\mu}} \mu d\mu + 2\pi \int_{\tau_{\lambda}}^{\tau_{s\lambda}} \int_0^1 J_{\lambda}(\tau'_{\lambda}, \mu) e^{\frac{\tau_{\lambda}-\tau'_{\lambda}}{\mu}} d\mu d\tau'_{\lambda} \quad (8.40.1)$$

$$F_{\lambda}^{\downarrow}(\tau_{\lambda}) = -2\pi \int_{-1}^0 I_{\lambda}(0, \mu) e^{\frac{\tau_{\lambda}}{\mu}} \mu d\mu + 2\pi \int_0^{\tau_{\lambda}} \int_{-1}^0 J_{\lambda}(\tau'_{\lambda}, \mu) e^{\frac{\tau_{\lambda}-\tau'_{\lambda}}{\mu}} d\mu d\tau'_{\lambda}. \quad (8.40.2)$$

For LW radiation and in the absence of scattering, the source function is just $B_{\lambda}(T)$. Upwelling radiation at the surface, which is treated as black, is then given by

$$I_{\lambda}(\tau_{s\lambda}, \mu) = B_{\lambda}(T_s) \quad 0 < \mu \leq 1. \quad (8.41.1)$$

Likewise, downwelling LW radiation at TOA must vanish, so

$$I_{\lambda}(0, \mu) = 0 \quad -1 \leq \mu < 0. \quad (8.41.2)$$

The upwelling and downwelling fluxes of LW radiation may then be written

$$F_{\lambda}^{\uparrow}(\tau_{\lambda}) = 2\pi B_{\lambda}(T_s) \int_0^1 e^{\frac{\tau_{\lambda}-\tau_{s\lambda}}{\mu}} \mu d\mu + 2 \int_{\tau_{\lambda}}^{\tau_{s\lambda}} \int_0^1 \pi B_{\lambda}[T(\tau'_{\lambda})] e^{\frac{\tau_{\lambda}-\tau'_{\lambda}}{\mu}} d\mu d\tau'_{\lambda} \quad (8.42.1)$$

$$F_{\lambda}^{\downarrow}(\tau_{\lambda}) = 2 \int_0^{\tau_{\lambda}} \int_0^1 \pi B_{\lambda}[T(\tau'_{\lambda})] e^{\frac{\tau'_{\lambda}-\tau_{\lambda}}{\mu}} d\mu d\tau'_{\lambda}. \quad (8.42.2)$$

The integrations over zenith angle may be expressed in terms of the exponential integral (Abramowitz and Stegun, 1972)

$$E_n(\tau) = \int_1^{\infty} \frac{e^{-x\tau}}{x^n} dx, \quad (8.43.1)$$

which satisfies

$$\frac{dE_n}{d\tau} = -E_{n-1}(\tau). \quad (8.43.2)$$

Letting $x = \mu^{-1}$ transforms (8.42) into

$$F_{\lambda}^{\uparrow}(\tau_{\lambda}) = 2\pi B_{\lambda}(T_s) E_3(\tau_{s\lambda} - \tau_{\lambda}) + 2 \int_{\tau_{\lambda}}^{\tau_{s\lambda}} \pi B_{\lambda}[T(\tau'_{\lambda})] E_2(\tau_{\lambda} - \tau'_{\lambda}) d\tau'_{\lambda} \quad (8.44.1)$$

$$F_{\lambda}^{\downarrow}(\tau_{\lambda}) = 2 \int_0^{\tau_{\lambda}} \pi B_{\lambda}[T(\tau'_{\lambda})] E_2(\tau_{\lambda} - \tau'_{\lambda}) d\tau'_{\lambda}. \quad (8.44.2)$$

Then integrating over wavelength recovers the total upward and downward fluxes

$$F^{\uparrow}(\tau) = 2 \int_0^{\infty} \pi B_{\lambda}(T_s) E_3(\tau_{s\lambda} - \tau_{\lambda}) d\lambda + 2 \int_{\tau}^{\tau_s} \int_0^{\infty} \pi B_{\lambda}[T(\tau'_{\lambda})] E_2(\tau_{\lambda} - \tau'_{\lambda}) d\lambda d\tau' \quad (8.45.1)$$

$$F^{\downarrow}(\tau) = 2 \int_0^{\tau} \int_0^{\infty} \pi B_{\lambda}[T(\tau'_{\lambda})] E_2(\tau_{\lambda} - \tau'_{\lambda}) d\lambda d\tau', \quad (8.45.2)$$

which describe LW radiative transfer in a non-scattering atmosphere.

8.4.1 Transmission function

Implementing (8.45) is complicated by the rapid variation with wavenumber of absorption cross section, which involves thousands of vibrational and rotational lines in the IR. In place of line-by-line calculations, radiative transfer is calculated

with the aid of a band-averaged transmission function that embodies the gross characteristics of the absorption spectrum in particular ranges of wavelength.

Averaging (8.44) over a frequency interval $\Delta\nu$, where $\nu = c\lambda^{-1}$ is symbolic of wavenumber, yields the band-averaged fluxes

$$F_{\bar{\nu}}^{\uparrow}(\tau_{\bar{\nu}}) = 2\pi B_{\bar{\nu}}(T_s) \int_{\Delta\nu} E_3(\tau_{s\nu} - \tau_{\nu}) \frac{d\nu}{\Delta\nu} + 2 \int_{\tau_{\nu}}^{\tau_{s\nu}} \pi B_{\bar{\nu}}[T(\tau'_{\nu})] \int_{\Delta\nu} E_2(\tau'_{\nu} - \tau_{\nu}) \frac{d\nu}{\Delta\nu} d\tau'_{\nu} \quad (8.46.1)$$

$$F_{\bar{\nu}}^{\downarrow}(\tau_{\bar{\nu}}) = 2 \int_0^{\tau_{\nu}} \pi B_{\bar{\nu}}[T(\tau'_{\nu})] \int_{\Delta\nu} E_2(\tau_{\nu} - \tau'_{\nu}) \frac{d\nu}{\Delta\nu} d\tau'_{\nu}. \quad (8.46.2)$$

In (8.46), $B_{\bar{\nu}}$ may be used for the band-averaged emission because the Planck function varies smoothly with frequency. The *band transmissivity* or *transmission function* is defined as

$$\mathcal{T}_{\bar{\nu}}(\tau_{\bar{\nu}}, \mu) = \frac{1}{\Delta\nu} \int_{\Delta\nu} e^{-\frac{\tau_{\nu}}{\mu}} d\nu. \quad (8.47)$$

Decreasing downward, $\mathcal{T}_{\bar{\nu}}$ represents the fractional intensity in the band and in direction μ that reaches optical depth $\tau_{\bar{\nu}}$. The *band absorptivity* is then

$$a_{\bar{\nu}} = 1 - \mathcal{T}_{\bar{\nu}}, \quad (8.48)$$

analogous to (8.11).

Averaging (8.47) over zenith angle defines the *diffuse flux transmission function* for the band

$$\begin{aligned} \mathcal{T}_{\bar{\nu}}(\tau_{\bar{\nu}}) &= \frac{\int_0^1 \mathcal{T}_{\bar{\nu}}(\tau_{\bar{\nu}}, \mu) \mu d\mu}{\int_0^1 \mu d\mu} \\ &= 2 \int_0^1 \mathcal{T}_{\bar{\nu}}(\tau_{\bar{\nu}}, \mu) \mu d\mu. \end{aligned} \quad (8.49)$$

It may be expressed in terms of the exponential integral as

$$\mathcal{T}_{\bar{\nu}}(\tau_{\bar{\nu}}) = 2 \int_{\Delta\nu} E_3(\tau_{\nu}) \frac{d\nu}{\Delta\nu}. \quad (8.50)$$

With the aid of (8.43.2), the band-averaged fluxes may then be expressed in terms of the flux transmission function

$$F_{\bar{\nu}}^{\uparrow}(\tau_{\bar{\nu}}) = \pi B_{\bar{\nu}}(T_s) \mathcal{T}_{\bar{\nu}}(\tau_{s\bar{\nu}} - \tau_{\bar{\nu}}) - \int_{\tau_{\nu}}^{\tau_{s\nu}} \pi B_{\bar{\nu}}[T(\tau'_{\nu})] \frac{d\mathcal{T}_{\bar{\nu}}(\tau'_{\nu} - \tau_{\bar{\nu}})}{d\tau'_{\nu}} d\tau'_{\nu} \quad (8.51.1)$$

$$F_{\bar{\nu}}^{\downarrow}(\tau_{\bar{\nu}}) = \int_0^{\tau_{\nu}} \pi B_{\bar{\nu}}[T(\tau'_{\nu})] \frac{d\mathcal{T}_{\bar{\nu}}(\tau_{\bar{\nu}} - \tau'_{\nu})}{d\tau'_{\nu}} d\tau'_{\nu}. \quad (8.51.2)$$

In principle, absorption characteristics of the band $\Delta\nu$ determine the corresponding transmission function and the vertical fluxes through (8.51). Collecting contributions from all absorbing bands then obtains the total upwelling and downwelling fluxes. However, in practice, even individual absorption bands are so complex as to make direct calculations impractical (cf. Figs. 8.11, 8.13). Instead, the transmission function is evaluated with the aid of band models that capture the salient features of the absorption spectrum in terms of properties such as mean line strength, line spacing, and line width. The regular band model of Elsasser (1938) treats a series of evenly spaced Lorentz lines, such as those comprising the 15- μm

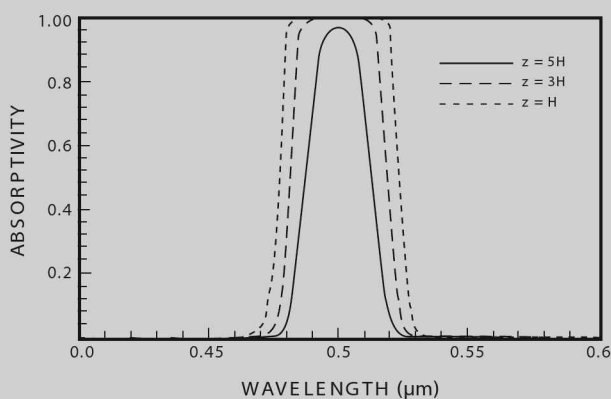


Figure 8.18 Absorptivity contributed by an individual absorption line at $0.5 \mu\text{m}$ for an absorber with a uniform mixing ratio and for the layer extending from the top of the atmosphere to a height z of 5 scale heights, 3 scale heights, and 1 scale height. As the layer extends downward, where absorber concentration increases exponentially, the line becomes saturated, first at its center but eventually over a widening range of wavelength.

band of CO_2 (Fig. 8.11a). The Goody (1952) random model treats a spectrum of lines that are randomly spaced, as is typical of the $6.3\text{-}\mu\text{m}$ band of water vapor (Fig. 8.11b).

The transmission function, although more complex than the exponential attenuation of monochromatic radiation (8.9), is far simpler than a line-by-line calculation over the band. It is instructive to consider how the transmission function, or alternatively the absorptivity (8.48), behaves with optical depth for an individual absorption line. Below a certain height, the absorptivity of a line may be represented in terms of the Lorentz profile (8.31.1). As τ_ν increases, so does a_ν (Fig. 8.18). Eventually, the absorptivity reaches unity near the line center. At that point, the corresponding frequencies are fully absorbed. The line is then *saturated*. This represents the limit of *strong absorption*.

Subsequent absorption can occur only in the wings of the line, which are pushed out by the ever widening region of saturation. In a band containing many lines, saturation leads to the absorption spectrum filling in between discrete features. This occurs first for the narrowly separated vibration-rotation transitions (e.g., Figs. 8.11, 8.13). Over sufficient optical depth, those clusters merge to form continuous bands of absorption, as appear in Figs. 8.1 and 8.9. The strong-absorption limit applies to bands that dominate vertical energy transfer in the atmosphere. It is the basis for a powerful approximation that reduces the 3-dimensional description of diffuse LW radiation to a 1-dimensional description.

8.4.2 Two-stream approximation

Embodied in the flux transmission function (8.50) is an integral over zenith angle, which collects contributions from diffuse radiation, and another integral over

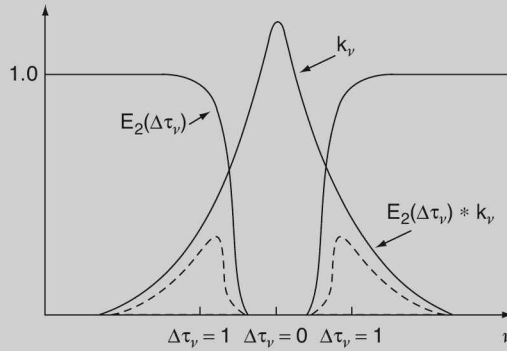


Figure 8.19 Factors $E_2(\Delta\tau_\nu)$ and k_ν appearing in the integral over zenith angle that is implicit to the band-averaged upwelling and downwelling fluxes (8.51). Due to their reciprocal nature, the factors' product is concentrated at frequencies corresponding to an optical depth of $\Delta\tau_\nu \cong 1$, which is associated with an effective zenith angle of $\bar{\mu}^{-1} \cong \frac{5}{3}$. Adapted from Andrews et al. (1987).

frequency. Because $E_3(\tau_\nu)$ depends on frequency, the two integrations are related. In the strong-absorption limit, the integral over zenith angle can be approximated fairly accurately by incorporating the spectral character of \mathcal{T}_ν .

Integrals on the right hand sides of (8.51) involve the term

$$\frac{d\mathcal{T}_\nu}{d\tau'_\nu}(\Delta\tau_\nu)d\tau'_\nu = 2 \int_{\Delta\nu} E_2(\Delta\tau_\nu)\rho k_\nu dz' \frac{d\nu}{\Delta\nu}, \quad (8.52.1)$$

where

$$\Delta\tau_\nu = |\tau'_\nu - \tau_\nu|. \quad (8.52.2)$$

If the band described by \mathcal{T}_ν is saturated, the factors k_ν and $E_2(\Delta\tau_\nu)$ have reciprocal behavior in frequency. As depicted in Fig. 8.19, $E_2(\Delta\tau_\nu)$ vanishes near the center of the band, for which the medium is optically thick. Conversely, it approaches unity in the wings of the band, for which the medium is optically thin. A transition between these extremes occurs at frequencies corresponding to an optical thickness of $\Delta\tau_\nu \cong 1$. The behavior of k_ν is just reversed. k_ν is large near the center of the band, but vanishes in the wings.

Due to the reciprocal nature of $E_2(\Delta\tau_\nu)$ and k_ν , their product in (8.52.1) is concentrated about an optical thickness of $\Delta\tau_\nu \cong 1$. Consequently, most of the energy exchange between two levels in (8.51) occurs along paths that have an optical thickness of unity. Along paths of optical thickness much greater than 1, the medium is optically thick. Radiation is then absorbed before it can traverse the levels, which therefore experience little interaction. Along paths of optical thickness much smaller than 1, the medium is optically thin. Radiation then traverses the levels with little attenuation, so the levels again experience little interaction.

For any specified $\Delta\tau_\nu$, an effective inclination $\bar{\mu}$ may be found such that the exponential integral over zenith angle is given by

$$2E_3(\Delta\tau_\nu) = e^{-\frac{\Delta\tau_\nu}{\bar{\mu}}}. \quad (8.53.1)$$

The parameter $\bar{\mu}^{-1}$ is called the *diffusivity factor*. As contributions to (8.51) are restricted to $\Delta\tau_v \cong 1$, it too is restricted to a narrow range of values. For $\Delta\tau_v = 1$, the diffusivity factor assumes the value

$$\bar{\mu}^{-1} \cong \frac{5}{3}. \quad (8.53.2)$$

The implication of this analysis is that the zenith angle dependence in (8.51) may be eliminated in favor of an effective inclination $\bar{\mu}$. The latter reduces the multi-dimensional description to integrals over optical depth alone. Diffuse transmission which is embodied in \mathcal{T}_v is then equivalent to that of a collimated beam inclined at a zenith angle of 53°

$$\mathcal{T}_v(\tau_v, \mu) \cong \mathcal{T}_v(\tau_v, \bar{\mu}). \quad (8.53.3)$$

Alternatively, it is equivalent to a beam inclined at zero zenith angle but through an optical thickness that is expanded by a factor of $\frac{5}{3}$.

Known as the *exponential kernel approximation*, this simplification follows from the fact that radiative exchange in strong bands is dominated by spectral intervals in which $\Delta\tau_v \cong 1$. For this reason, most of the LW radiation emitted by the Earth's surface is captured in the lower troposphere. The latter is rendered optically thick by strong absorption bands of water vapor and carbon dioxide. Likewise, incident SW radiation in particular wavelengths of UV is absorbed in the stratosphere over a limited range of altitude (Fig. 8.3). The latter is rendered optically thick by photodissociation at those wavelengths. The development leading to (8.53) is one of several so-called *two stream approximations* that eliminate the zenith angle dependence in F^\uparrow and F^\downarrow . Applicable under fairly general circumstances (even in the presence of scattering), this formalism leads to diffusivity factors in the range $\frac{3}{2} \leq \bar{\mu}^{-1} \leq 2$, depending on the particular approximation adopted.

The full description of radiative transfer in a plane parallel atmosphere may be reduced to a vertical description by taking $\mu = \pm\bar{\mu}$ for upwelling and downwelling radiation and introducing the transformation

$$\tau_\lambda^* = \bar{\mu}^{-1} \tau_\lambda \quad (8.54.1)$$

$$F_\lambda^{\uparrow\downarrow}(\tau_\lambda^*) = \pi I_\lambda(\tau_\lambda, \pm\bar{\mu}) \quad (8.54.2)$$

$$J_\lambda^*(\tau_\lambda^*) = \pi J_\lambda(\tau_\lambda, \pm\bar{\mu}), \quad (8.54.3)$$

which relies on hemispheric isotropy. Then integrating the radiative transfer equation (8.37) over upward and downward half-spaces yields for the budgets of upwelling and downwelling radiation

$$\frac{dF_\lambda^\uparrow}{d\tau_\lambda^*} = F_\lambda^\uparrow - J_\lambda^* \quad (8.55.1)$$

$$-\frac{dF_\lambda^\downarrow}{d\tau_\lambda^*} = F_\lambda^\downarrow - J_\lambda^*. \quad (8.55.2)$$

The foregoing development applies to LW radiation, which is inherently diffuse. SW radiation, which is parallel-beam in the absence of scattering, involves only a single direction μ .

8.5 THERMAL EQUILIBRIUM

8.5.1 Radiative equilibrium in a gray atmosphere

We are now in a position to evaluate the thermal structure toward which radiative transfer drives the atmosphere. To do so, we consider a simple model of the Earth's atmosphere: a *gray atmosphere* is transparent to SW radiation but absorbs LW radiation with a constant absorption cross section that is independent of wavelength, temperature, and pressure.

Consider a plane parallel gray atmosphere that is nonscattering, motionless, and in thermal equilibrium with incoming SW radiation and with a black underlying surface. Individual layers of the atmosphere then interact only through LW radiation.

For an incremental layer of thickness dz , the First Law (2.22.2) implies

$$\begin{aligned} \rho c_p \frac{dT}{dt} - \frac{dp}{dt} &= \rho \dot{q} \\ &= -\frac{dF}{dz}, \end{aligned} \quad (8.56.1)$$

where

$$F = F^\uparrow - F^\downarrow \quad (8.56.2)$$

is the net LW flux integrated over wavelength and \dot{q} is the local rate heat is absorbed per unit mass or the *specific heating rate*. The convergence of LW flux on the right hand side of (8.56.1) represents the local radiative heating rate per unit volume. It forces the atmosphere's thermal structure. In equilibrium, the left hand side vanishes. Then the net flux must be independent of height

$$F = \text{const.} \quad (8.57)$$

This implies that radiative energy does not accumulate within individual layers.

The upwelling and downwelling components of F are governed by (8.55). Adding and subtracting yields the equivalent system

$$\frac{dF}{d\tau^*} = \bar{F} - 2B^* \quad (8.58.1)$$

$$\frac{d\bar{F}}{d\tau^*} = F, \quad (8.58.2)$$

where

$$\bar{F} = F^\uparrow + F^\downarrow \quad (8.58.3)$$

represents the total flux emanating from an incremental layer and all quantities have been integrated over wavelength. With (8.57), (8.58.2) gives

$$\bar{F} = F\tau^* + c, \quad (8.59.1)$$

whereas (8.58.1) reduces to

$$\bar{F} = 2B^*. \quad (8.59.2)$$

Emission is then described by

$$B^*(\tau^*) = \frac{F}{2}\tau^* + B_0^*. \quad (8.60)$$

It increases linearly with optical depth, from its value B_0^* at TOA (Fig. 8.20).

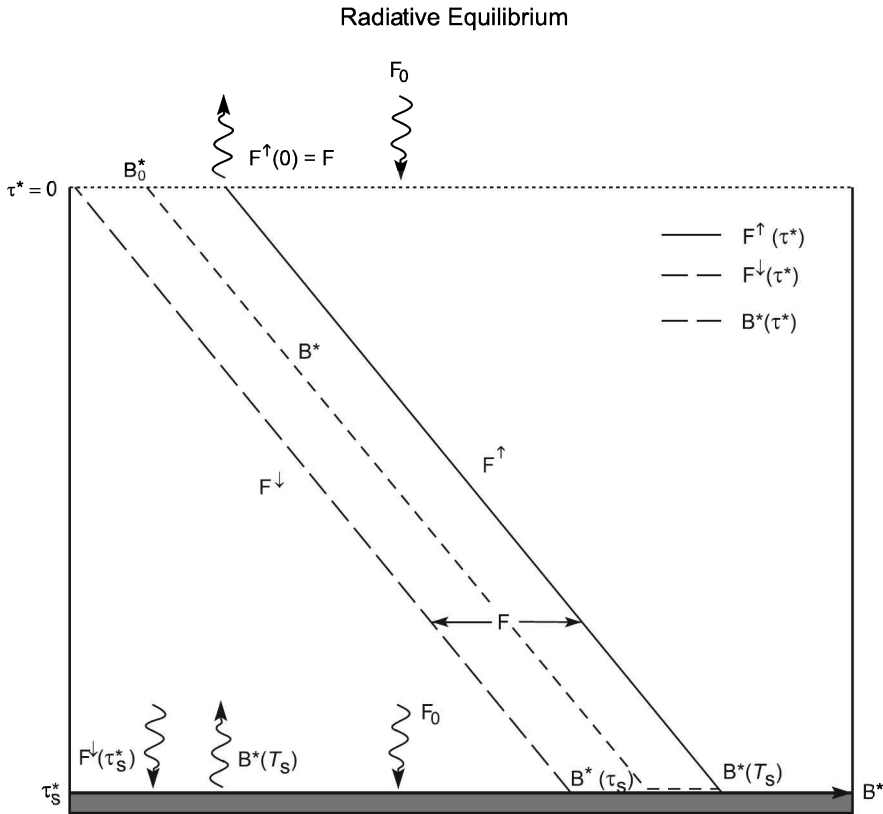


Figure 8.20 Upwelling and downwelling LW fluxes and LW emission in a gray atmosphere that is in radiative equilibrium with an incident SW flux F_0 and a black underlying surface. Note: the emission profile is discontinuous at the surface.

For thermal equilibrium of the Earth-atmosphere system, the incident SW flux

$$F_0 = (1 - \mathcal{A})\bar{F}_s \tag{8.61}$$

(Sec. 1.4.1) must be balanced by the outgoing LW flux at the top of the atmosphere

$$F^\uparrow(0) = F_0. \tag{8.62}$$

Because the downwelling LW flux at TOA vanishes, $F(0)$ and $\bar{F}(0)$ each reduce to $F^\uparrow(0)$. Then, by (8.57),

$$\begin{aligned} F &= F_0. \\ &= \text{const.} \end{aligned} \tag{8.63}$$

The net LW flux at any level equals the SW flux at TOA. Incorporating (8.59.2) then yields

$$B^*(\tau^*) = \frac{F_0}{2}(\tau^* + 1), \tag{8.64}$$

which is plotted in Fig. 8.20 along with F^\uparrow and F^\downarrow . Under radiative equilibrium, the upwelling and downwelling fluxes differ by a constant. Net heating at any level therefore vanishes (8.56.1).

Consider now thermal equilibrium of the surface. Because it is black, the surface fully absorbs the incident SW flux. It also absorbs the downwelling LW flux emitted by the atmosphere $F^\downarrow(\tau_s^*)$. These must be balanced by the upwelling LW flux emitted by the surface:

$$B^*(T_s) = F_0 + F^\downarrow(\tau_s^*). \quad (8.65)$$

Subtracting (8.59.2) and (8.63) gives the downwelling LW flux emitted by the atmosphere

$$F^\downarrow(\tau_s^*) = B^*(\tau_s^*) - \frac{F_0}{2}. \quad (8.66)$$

Combining it with (8.65) yields

$$B^*(T_s) = B^*(\tau_s^*) + \frac{F_0}{2}. \quad (8.67)$$

According to (8.67), the temperature under radiative equilibrium is discontinuous at the surface: The ground, with emission $B^*(T_s)$, is warmer than the overlying air, with emission $B^*(\tau_s^*)$.

The emission profile (8.64) determines the radiative-equilibrium temperature through

$$\begin{aligned} B^*(\tau^*) &= \pi B[T(\tau^*)] \\ &= \sigma T^4. \end{aligned} \quad (8.68)$$

Because the atmosphere is hydrostatic, density decreases with height exponentially (1.17). By (8.36), so does τ^* . Further, most of the atmosphere's opacity follows from water vapor and cloud (Fig. 8.1), which are concentrated in the lowest levels. Therefore, the linear variation with optical depth of B^* (8.64) translates into a steep decrease with height of temperature. At the surface, the lapse rate is infinite because, there, the radiative-equilibrium temperature is discontinuous.

Optical depth is taken to vary with height exponentially

$$\tau = \tau_s e^{-\frac{z}{h}}, \quad (8.69)$$

where $h = 2$ km is symbolic of the absolute concentration of water vapor (Fig. 1.17). Then $F_0 = 240 \text{ W m}^{-2}$ obtains the temperature profiles shown in Fig. 8.21 (solid). For given $\tau_s > 0$, the surface temperature exceeds that in the absence of an atmosphere ($\tau_s = 0$). It increases with the atmosphere's optical depth. This dependence on τ_s is a manifestation of the greenhouse effect. Radiative equilibrium temperature is maximum at the ground, decreasing upward steeply. It jumps discontinuously to the surface temperature T_s . As $z \rightarrow \infty$, the radiative-equilibrium temperature approaches a finite limiting value $\left(\frac{F_0}{2\sigma}\right)^{\frac{1}{4}} \cong 215^\circ\text{K}$, called the *skin temperature*.

In a neighborhood of the surface, temperature decreases upward sharply - more sharply, in fact, than the saturated adiabatic lapse rate $\Gamma_s \cong 5.5^\circ\text{K km}^{-1}$ (dotted). The latter corresponds to moist neutral stability. Thus, in the lowest levels, each of the radiative-equilibrium temperature profiles in Fig. 8.21 is statically unstable. For an optical depth of $\tau_s = 4$, the radiative-equilibrium profile is conditionally unstable below

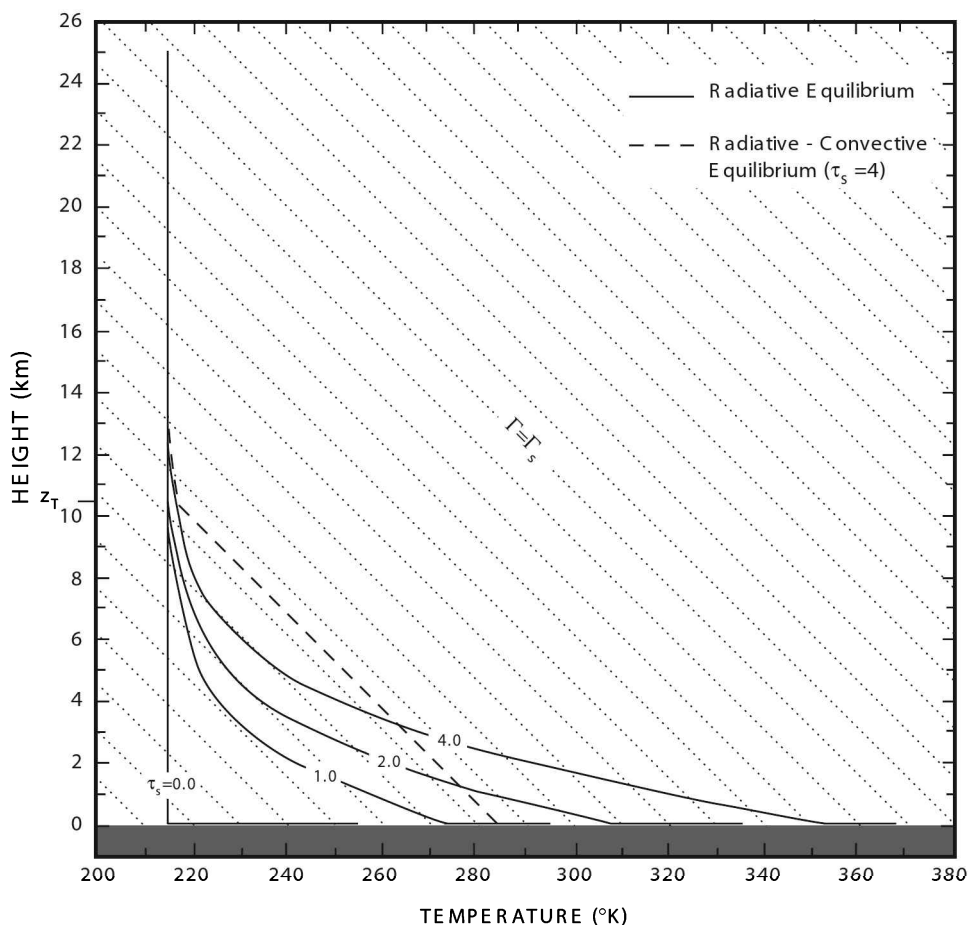


Figure 8.21 Radiative equilibrium temperature (solid) for the gray atmosphere in Fig. 8.20, with a profile of optical depth representative of water vapor (8.69). Presented for several atmospheric optical depths τ_s . Saturated adiabatic lapse rate (dotted) and radiative-convective equilibrium temperature for $\tau_s = 4$ (dashed) superposed.

6 km, and absolutely unstable below about 5 km. At the ground, it has a lapse rate of $\sim 36^\circ\text{K km}^{-1}$.

8.5.2 Radiative-convective equilibrium

If air motion is accounted for, the conditions produced under radiative equilibrium cannot be maintained. Convection develops spontaneously, neutralizing the unstable stratification that was introduced by radiative transfer. It leads to *radiative-convective equilibrium*. Superposed in Fig. 8.21 for $\tau_s = 4$ (dashed), radiative-convective equilibrium produces two layers of distinctly different structure. Below a height z_T , thermal structure is controlled by convective overturning that is driven by radiative heating and its destabilization of the stratification. This layer constitutes the troposphere

in this framework. Because convective overturning operates on a time scale much shorter than radiative transfer, it drives stratification below z_T to neutral stability. Under saturated conditions, this corresponds to a lapse rate of Γ_s (Sec. 7.6.1). Heat supplied at the Earth's surface is transferred upward and mixed vertically over the convective layer. This process maintains a uniform profile of equivalent potential temperature, $\theta_e \cong \text{const}$. In the layer above z_T , thermal structure remains close to radiative equilibrium because, at those levels, radiative transfer stabilizes the stratification. This layer constitutes the stratosphere in this framework. Tropospheric air displaced upward by convection cools along the saturated adiabat. It therefore penetrates little above the tropopause height z_T , where the saturated-adiabatic and radiative-equilibrium profiles cross.

The particular form of radiative-convective equilibrium assumed depends on the parameterization of cloud and convective heat flux at the surface, which dictate the height of the tropopause. For instance, taking the height of the convective layer to equal the maximum height of instability under radiative equilibrium predicts one tropopause height. Taking the air temperature at $z = 0$ to equal the surface temperature under radiative equilibrium leads to a different tropopause height.

An alternative that circumvents ambiguities surrounding convection requires the convective layer to supply the same upward radiation flux at the tropopause as would be supplied under radiative equilibrium (Goody and Yung, 1995). This formalism is tantamount to presuming that the stratosphere is unaffected by vertical motion below. It provides a self-consistent rationale for determining radiative-convective equilibrium. Integrating (8.55) as in the development of (8.29) leads to expressions valid under arbitrary conditions for the upward and downward radiation fluxes:

$$F^\uparrow(\tau^*) = B(T_s)e^{\tau^* - \tau_s^*} - \int_{\tau_s^*}^{\tau^*} e^{\tau^* - \tau'} B^*[T(\tau')] d\tau' \quad (8.70.1)$$

$$F^\downarrow(\tau^*) = \int_0^{\tau^*} e^{\tau' - \tau^*} B(\tau') d\tau', \quad (8.70.2)$$

where convection has been presumed efficient enough to maintain the surface at the same temperature as the overlying air. Then equating (8.70.1) to the upward flux under radiative equilibrium gives

$$F^\uparrow(\tau^*) = \frac{F_0}{2}(\tau^* + 2), \quad (8.71)$$

which follows in the same manner as (8.66). For $\tau_s = 4$, this treatment yields a surface temperature of $T_s \cong 285^\circ\text{K}$, close to the observed global-mean surface temperature of the Earth. Superposed in Fig. 8.21 is the radiative-convective equilibrium temperature that results (dashed). Temperature under radiative-convective equilibrium is cooler in the lower troposphere and warmer in the upper troposphere than results under radiative equilibrium. Stability in the lowest levels has been neutralized. Radiative-convective equilibrium achieves a tropopause at about 10.5 km, in qualitative agreement with observed thermal structure. Relative to the characteristic extinction scale, $h = 2$ km, z_T lies several optical depths above the Earth's surface. Temperature in the stratosphere is therefore close to the atmosphere's skin temperature.

The corresponding upward and downward LW fluxes are plotted as functions of height in Fig. 8.22a (dashed), along with those under radiative equilibrium (solid). Reduced temperature in the lower troposphere leads to smaller upward and downward

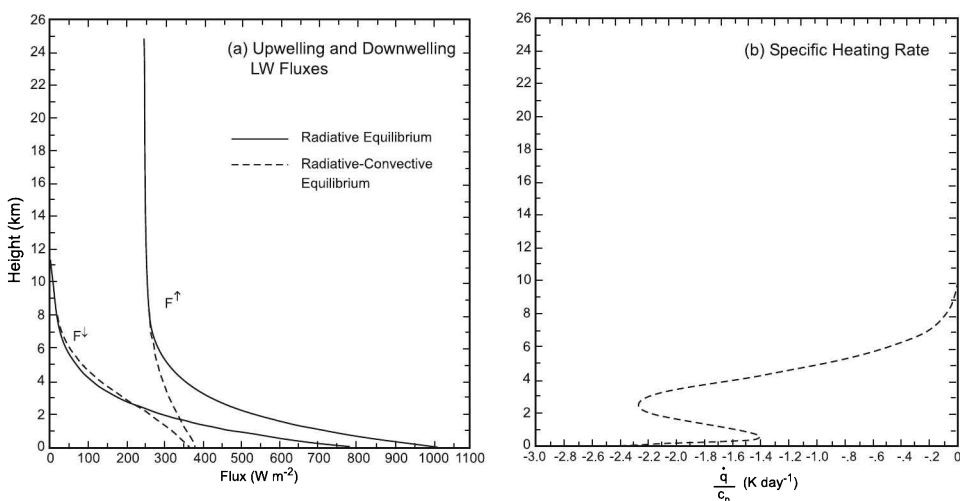


Figure 8.22 (a) Upwelling and downwelling fluxes as functions of height in the gray atmosphere in Fig. 8.20 for radiative equilibrium (solid) and radiative-convective equilibrium (dashed). (b) Specific heating rate under radiative-convective equilibrium.

fluxes. They merge with the radiative-equilibrium profiles above z_T . To satisfy overall equilibrium, F^\downarrow approaches 0, whereas F^\uparrow approaches 240 W m^{-2} as $z \rightarrow \infty$. Although the radiative-equilibrium fluxes preserve a constant difference, the net flux under radiative-convective equilibrium varies with altitude below z_T . Implied is nonzero radiative heating inside the troposphere (8.56).

The specific heating rate $\frac{\dot{q}}{c_p}$, which has dimensions of degrees per day, is plotted in Fig. 8.22b. It is negative throughout the troposphere, which experiences net radiative cooling (Fig. 1.32). Radiative cooling approaches zero inside the stratosphere, where thermal structure comes under radiative equilibrium. Inside the troposphere, radiative cooling must, at each level, be balanced by convective heating. The latter is supplied by upward heat transfer from the Earth's surface. This is consistent with observed thermal structure inside the tropical troposphere, where deep convection prevails (Sec. 7.7). Notice that radiative cooling is greatest at (1) the surface, where temperature is a maximum, and (2) the top of the extinction layer ($z \cong 2\text{--}4 \text{ km}$), corresponding to water vapor, above which LW emission passes freely to space. Reaching a maximum of about 2°K day^{-1} , the cooling rate in Fig. 8.22b is in qualitative agreement with detailed calculations of radiative cooling based on observed behavior (cf. Fig. 8.24a).

More sophisticated calculations of radiative-convective equilibrium incorporate distributions of optically active species, mean cloud cover, and surface absorption of SW radiation. Figure 8.23 shows the results of a calculation that accounts for radiative transfer by water vapor, carbon dioxide, and ozone. The radiative-equilibrium temperature (solid) decreases steeply from about 340°K at the surface to a tropopause value of about 185°K near 10 km. Most of the troposphere is absolutely unstable. In the stratosphere, temperature increases upward due to absorption of SW radiation by ozone. Radiative-convective equilibrium (dashed) produces thermal structure which resembles that of the gray atmosphere. Relative to thermal structure under radiative

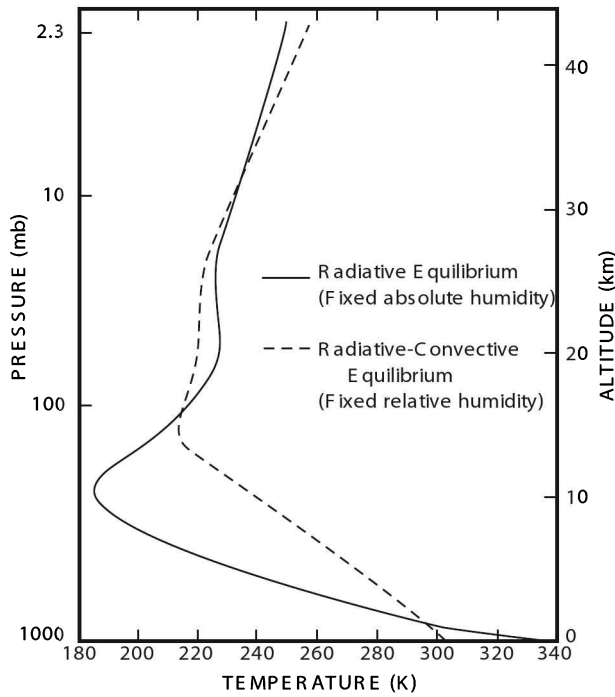


Figure 8.23 Temperature under radiative equilibrium (solid) and radiative-convective equilibrium (dashed) from calculations that include mean distributions of water vapor, carbon dioxide, and ozone. Adapted from Manabe and Wetherald (1967).

equilibrium, temperature is cooler in the lower troposphere and warmer in the upper troposphere.

8.5.3 Radiative heating

Knowledge of the mean distributions of temperature and optically active constituents enables LW heating to be inferred via (8.56) from the upwelling and downwelling fluxes. For an individual band, the specific heating rate may be expressed

$$\begin{aligned} \frac{\dot{q}}{c_p} &= -\frac{1}{\rho c_p} \frac{dF}{dz^*} \\ &= \frac{k}{c_p} \frac{dF}{d\tau^*}, \end{aligned} \quad (8.72)$$

where (8.53) has been incorporated into height and the frequency dependence is implicit. Differentiating and subtracting (8.51) gives

$$\begin{aligned} \frac{dF}{d\tau^*} &= B^*(T_s) \frac{dT(\tau_s^* - \tau^*)}{d\tau^*} - \int_{\tau^*}^{\tau_s^*} B^*[T(\tau')] \frac{d^2\mathcal{T}(\tau' - \tau^*)}{d\tau' d\tau^*} d\tau' \\ &\quad - \int_0^{\tau^*} B^*[T(\tau')] \frac{d^2\mathcal{T}(\tau^* - \tau')}{d\tau' d\tau^*} d\tau' + B^*[T(\tau^*)] \left. \frac{d\mathcal{T}(|\tau' - \tau^*|)}{d\tau'} \right|_{\tau'=\tau_-}^{\tau'=\tau_+}. \end{aligned}$$

The specific heating rate can then be arranged into the form

$$\begin{aligned} \frac{\dot{q}(\tau^*)}{c_p} = \frac{k(\tau^*)}{c_p} \left\{ B^*[T(\tau^*)] \frac{dT(\tau^* - 0)}{d\tau^*} + \left[B^*(T_s) - B^*(\tau^*) \right] \frac{dT(\tau_s^* - \tau^*)}{d\tau^*} \right. \\ \left. - \int_{\tau^*}^{\tau_s^*} \left[B^*(\tau') - B^*(\tau^*) \right] \frac{d^2T(\tau' - \tau^*)}{d\tau^* d\tau'} d\tau' \right. \\ \left. - \int_0^{\tau^*} \left[B^*(\tau') - B^*(\tau^*) \right] \frac{d^2T(\tau^* - \tau')}{d\tau^* d\tau'} d\tau' \right\}. \end{aligned} \quad (8.73)$$

Equation (8.73) collects contributions to $\frac{\dot{q}}{c_p}$ from LW interactions between the level τ^* and its environment. $\frac{dT(\tau^* - \tau')}{d\tau'}$ then serves as an influence function between that level and another τ' . The first term on the right hand side describes interaction with space. Because $\frac{dT}{d\tau} < 0$ (8.47), this term is always negative. It represents cooling to space. The second term, which describes interaction with the surface, is negative when the level τ^* is warmer than the ground. The last two terms are *exchange integrals* that collect contributions from levels τ' above and below τ^* . Because $\frac{d^2T}{d\tau d\tau'} < 0$, those contributions represent cooling when the temperature at level τ^* exceeds that at level τ' .

Figure 8.24a illustrates the dominant contributions to LW heating. Except for ozone, the primary LW absorbers cool the atmosphere. Water vapor dominates LW cooling in the troposphere. Accounting for 80–90% of overall cooling, it leads to a globally averaged cooling rate of about 2°K day^{-1} . In the stratosphere, the $15\text{-}\mu\text{m}$ band of CO_2 dominates LW cooling. Together with the $9.6\text{-}\mu\text{m}$ band of O_3 , it produces a maximum cooling rate near the stratopause of $\sim 12^\circ\text{K day}^{-1}$. Consistent with the dominance of CO_2 cooling in the stratosphere is a cooling trend observed at those altitudes, which accompanies increasing CO_2 and a warming trend near the Earth's surface (Fig. 1.42). Because the $9.6\text{-}\mu\text{m}$ band of ozone lies within the atmospheric window, it also produces heating in the lower stratosphere, by absorbing upwelling LW radiation from below.³

In contrast to LW radiation, SW radiation produces only heating, as the atmosphere does not emit at those wavelengths. For a particular wavelength, the specific heating rate is given by

$$\dot{q} = -k\mu_s \frac{dF}{d\tau},$$

where $\mu_s = -\mu > 0$ refers to the *solar zenith angle* and, as above, the wavelength dependence is implicit. Incorporating (8.7) transforms this into

$$\begin{aligned} \dot{q} &= -k\mu_s \frac{d}{d\tau} \left[I_0 e^{-\frac{\tau}{\mu_s}} \right] \\ &= kI_0 e^{-\frac{\tau}{\mu_s}}. \end{aligned} \quad (8.74)$$

³ Even though specific heating rates in the stratosphere are larger than those in the troposphere, volume heating rates $\rho\dot{q}$ are smaller by two orders of magnitude.

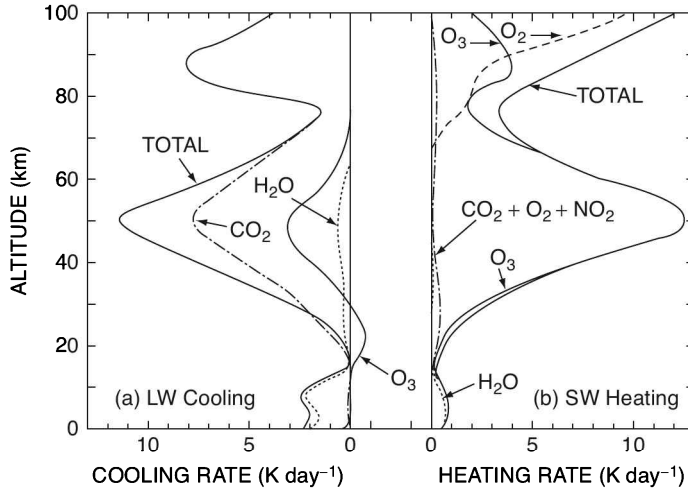


Figure 8.24 Global-mean profiles of (a) LW cooling and (b) SW heating. Contributions from individual radiatively active constituents also shown. After London (1980).

Then the volume heating rate follows as

$$\begin{aligned}\rho\dot{q} &= \rho k I_0 e^{-\frac{\tau}{\mu_s}} \\ &= n \hat{k} I_0 e^{-\frac{\tau}{\mu_s}},\end{aligned}\quad (8.75)$$

where ρ and n refer to absorbers at the wavelength under consideration.

Heating varies strongly with the absolute concentration n of absorber. Dominated by the variation of air density, the absorber concentration may be modeled as

$$\frac{n}{n_0} = e^{-\frac{z}{H}}.$$

Then the optical depth also varies exponentially and (8.75) becomes

$$\rho\dot{q} = n_0 \hat{k} I_0 e^{-\left[\tau_0 e^{-\frac{z}{H}} + \frac{z}{H}\right]},\quad (8.76.1)$$

where

$$\tau_0 = \frac{n_0 \hat{k} H}{\mu_s}.\quad (8.76.2)$$

The volume heating rate (8.76) possesses a single maximum at the height

$$\frac{z_0}{H} = \ln \tau_0.\quad (8.77)$$

It corresponds to a slant optical path from TOA of unity. Above that level, heating decreases exponentially with air density. Below, it decreases even faster due to attenuation in the first exponential term of (8.76.1). Therefore, SW heating is concentrated within about a scale height of the level z_0 . It defines the *Chapman layer* for this wavelength. Consequently, the penetration altitude in Fig. 8.3 for a particular wavelength also describes where most of the SW absorption and heating occur. As shown in Fig. 8.25, increased solar zenith angle reduces SW heating, because the vertical flux is

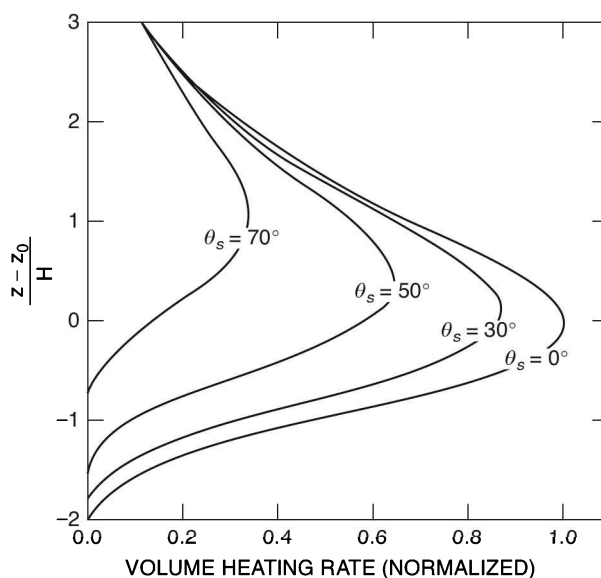


Figure 8.25 Volume heating rate due to absorption of SW radiation at solar zenith angle θ_s . Concentration of heating within a depth of order H defines the *Chapman layer* for a particular wavelength. After Banks and Kocharts (1973).

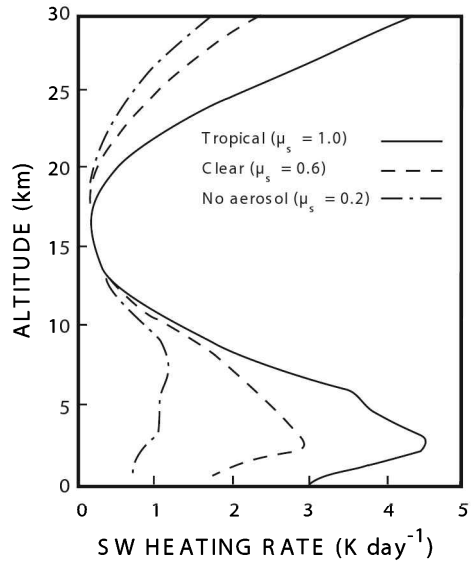
diminished according to μ_s . It also increases the altitude z_0 , because the optical path length traversed by solar radiation is elongated.

The extinction cross section k for different wavelengths results from different species (8.6). They, in turn, have different vertical profiles. Consequently, the full spectrum of SW radiation produces a series of Chapman layers, which follow from the vertical distributions of optically active species. According to Fig. 8.3, energetic radiation at wavelengths shorter than 200 nm is absorbed in the upper mesosphere and thermosphere by O_2 in the Schumann-Runge bands and continuum. Wavelengths of 200–300 nm are absorbed in the mesosphere and stratosphere by O_3 in the Hartley band. Longer wavelengths penetrate to the surface, with partial absorption in the visible by ozone and in the near IR by water vapor and carbon dioxide.

Ozone accounts for most of the SW absorption in the stratosphere and mesosphere (Fig. 8.24b). It produces maximum heating near the stratopause, even though the absolute concentration of ozone maximizes considerably lower (Fig. 1.20). In the global-mean, SW heating and LW cooling at these heights nearly cancel. The middle atmosphere is therefore close to radiative equilibrium. In the troposphere, water vapor is the dominant SW absorber. Achieving a global-mean heating rate of about 1°K day^{-1} , SW absorption does not completely offset LW cooling. Thus, radiative transfer leads to net cooling of the troposphere of order 1°K day^{-1} . It is compensated by convective heating, which derives from the upward transfer of heat from the Earth's surface (Fig. 1.32).

Heating rates depicted in Fig. 8.24 apply to a clear atmosphere under global-mean conditions. Relative to those, local conditions can deviate dramatically. Figure 8.26 shows that, under conditions typical of the tropical troposphere, SW heating rates can

Figure 8.26 SW heating rate under different atmospheric conditions relating to solar elevation and the presence of aerosol. After Liou (1980).



be substantially greater (see also Reynolds et al., 1975). Likewise, the radiative balance in the stratosphere suggested by Fig. 8.24 for the global mean does not apply locally. As shown in Fig. 8.27, the winter stratosphere experiences strong net cooling - more than $8^{\circ}\text{K day}^{-1}$ inside polar night. It implies this region is warmer than radiative equilibrium. As in the troposphere, net radiative cooling must be balanced by a mechanical transfer of energy. That energy transfer occurs through planetary waves

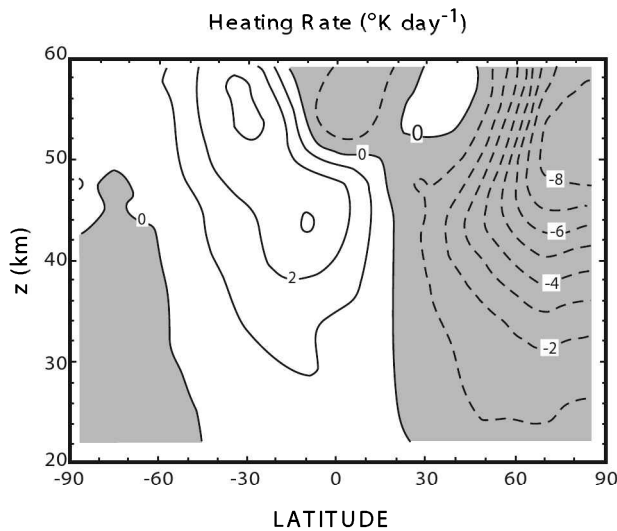


Figure 8.27 Net radiative heating in the middle atmosphere during January, as derived from Nimbus-7 LIMS. After Kiehl and Solomon (1986).

(Fig. 1.10). By driving the circulation out of radiative equilibrium, they perform work on the stratosphere. Manifest in temperature warmer than radiative equilibrium, that work is converted into heat and eventually rejected to space via LW cooling.

8.6 THERMAL RELAXATION

If a layer is in radiative equilibrium with its surroundings, LW and SW contributions to its radiation budget balance. Net heating then vanishes. If disturbed from radiative equilibrium (e.g., by motions that displace air into different radiative environments), a layer will experience net heating or cooling (8.73), which tends to restore the layer to radiative equilibrium. For example, polar air drawn equatorward by a cyclone will find itself colder than the radiative-equilibrium temperature of its new environment (Fig. 1.18). Net radiative heating then acts to destroy the temperature anomaly by driving that air toward the local radiative-equilibrium temperature. Because the circulation is related to thermal structure, damping anomalous temperature also damps anomalous motion.

In the absence of other forms of heat transfer, the temperature field is described by (8.56), with the heating rate given by (8.73). The first term in (8.73) describes cooling to space. It often dominates radiative heating. Figure 8.28 compares the total heating for water vapor, carbon dioxide, and ozone against the contributions from cooling to space alone. Over much of the troposphere and stratosphere, cooling to space provides an accurate approximation to net radiative heating. Exceptional is ozone in the lower stratosphere, which absorbs upwelling 9.6- μm radiation from below.

The agreement in Fig. 8.28 under widely varying circumstances is the basis for the *cool-to-space approximation*. For a particular frequency band, it is expressed by

$$\frac{dT}{dt} \cong -\frac{1}{\rho c_p} B^*[T(z^*)] \frac{dT_v(z^* - \infty)}{dz^*}, \quad (8.78)$$

where vertical motion (e.g., $\frac{dp}{dt}$) is ignored, z^* includes the diffusivity factor (8.54.1), and summation over different absorbers is understood. Away from the surface and in the absence of strong curvature of the emission profile, (8.78) accurately represents cooling due to both water vapor and carbon dioxide. These conditions hold for temperature disturbances of large vertical scale. Interactions with underlying and overlying layers are then negligible. Under these conditions, emission to space of monochromatic LW radiation takes place from a Chapman layer, analogous to absorption of SW radiation. Emission over a finite spectral band can thus be interpreted in terms of a series of Chapman layers.

Consider a disturbance $T'(z, t)$ to an equilibrium temperature profile $T_0(z)$.⁴ If $\frac{T'}{T_0} \ll 1$ and if the cool-to-space approximation applies, then emission in (8.78)

⁴ The equilibrium thermal structure may be maintained by heat transfer other than radiation alone. However, we will take T_0 to represent the radiative-equilibrium temperature.

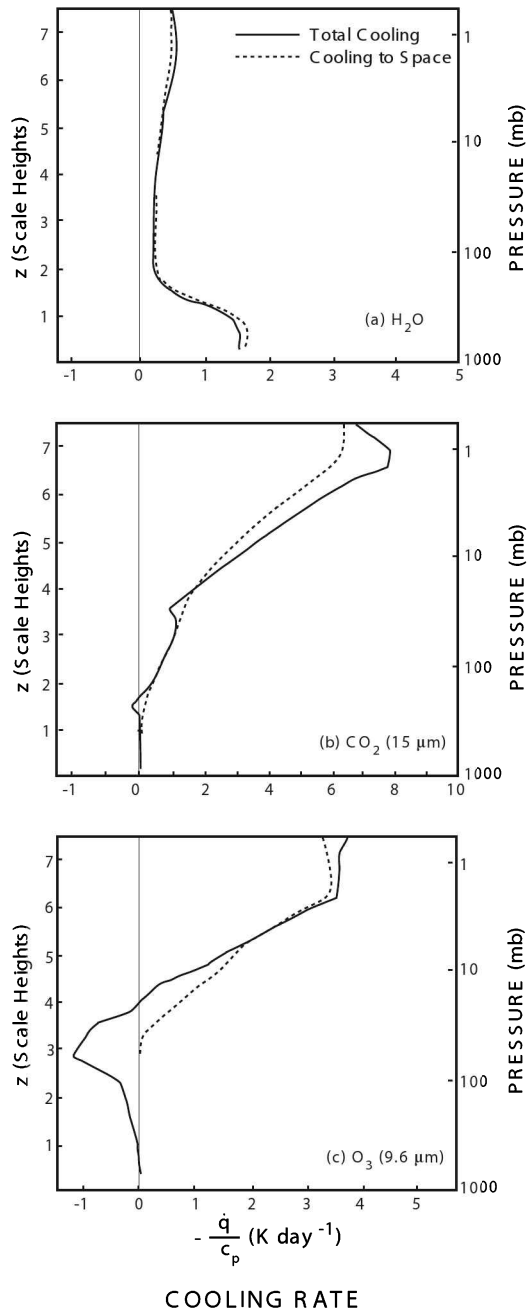


Figure 8.28 Comparison of the total cooling rate (solid) vs the contribution from cooling to space (dashed) for (a) water vapor, (b) carbon dioxide, and (c) ozone. Adapted from Rodgers and Walshaw (1966).

may be linearized in T' , yielding

$$\begin{aligned}\frac{\partial T'}{\partial t} &= -\frac{1}{\rho_0 c_p} \frac{dB^*}{dT} [T_0(z^*)] \frac{dT_v(z^* - \infty)}{dz^*} \cdot T' \\ &= -\alpha(z^*) T'.\end{aligned}\quad (8.79)$$

Known as the *Newtonian cooling approximation*, (8.79) governs the evolution of anomalous temperature. The Newtonian cooling coefficient $\alpha(z^*)$ varies spatially through the equilibrium thermal structure $T_0(z^*)$. This approximation is of great practical importance because it eliminates the rather cumbersome interactions in (8.73) in favor of a simple expression that depends only on the local temperature. It is also linear in disturbance temperature. Even though it breaks down for CO_2 in the mesosphere (Andrews et al., 1987), the Newtonian cooling approximation is widely adopted.

Under the influence of Newtonian cooling, a temperature disturbance relaxes exponentially toward radiative equilibrium. Owing to its linearity, (8.79) may be cast into the form of an e-folding time

$$\begin{aligned}-\frac{1}{T'} \frac{\partial T'}{\partial t} &= \frac{4\sigma T_0^3}{\rho_0 c_p} \frac{dT_v(z^* - \infty)}{dz^*} \\ &= t_{rad}^{-1}.\end{aligned}\quad (8.80)$$

t_{rad} measures the efficiency of radiative transfer in relation to other factors that influence thermal structure. Taking values representative of the troposphere (Prob. 8.17) yields a radiative time scale of order 10 days.

The time scale $t_{rad} = \alpha^{-1}$ is an order of magnitude longer than the characteristic time scale of air motion. Radiative transfer is therefore inefficient compared with dynamical influences. For this reason, air motion controls tropospheric properties like thermal structure and vertical stability. In the stratosphere, where water vapor and cloud are rare, cooling to space is stronger. There, t_{rad} decreases to only 3–5 days. Although shorter, this is still long enough to permit air motion to influence many stratospheric properties.

More accurate formulations of heat transfer apply to temperature disturbances whose vertical scales are not large enough to neglect exchange between neighboring layers. Plotted in Fig. 8.29 is the thermal damping rate for sinusoidal vertical structure with several vertical wavelengths λ_z . Thermal dissipation minimizes in the limit $\lambda_z \rightarrow \infty$. This limit corresponds to deep temperature structure that is governed by Newtonian cooling (8.79). Under those circumstances, t_{rad} decreases from 10–15 days in the troposphere to about 5 days near the stratopause. Smaller vertical scales experience significantly faster damping due to exchange with neighboring layers. Relaxation times of only a couple of days are then experienced in the upper stratosphere and mesosphere.

8.7 THE GREENHOUSE EFFECT

The radiative-equilibrium surface temperature T_s is significantly warmer than that in the absence of an atmosphere. Combining (8.67) and (8.64) for a gray atmosphere yields

$$B^*(T_s) = \frac{F_0}{2} (\tau_s^* + 2). \quad (8.81)$$

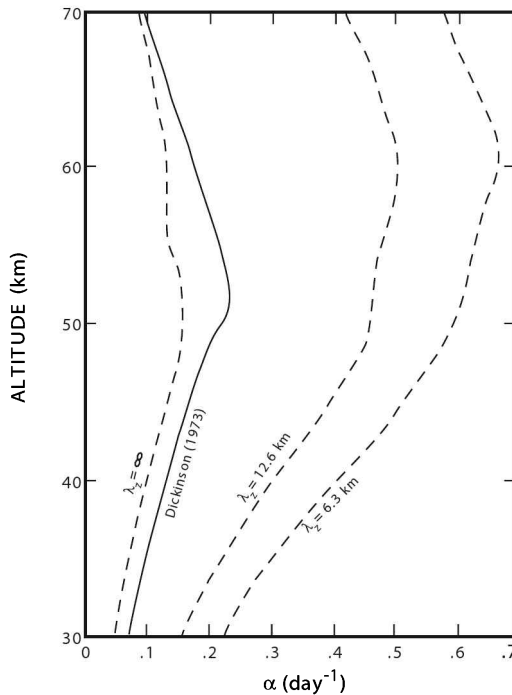


Figure 8.29 Thermal damping rate for temperature disturbances with sinusoidal vertical structure of vertical wavelength λ_z (dashed). After Fels (1982). Newtonian cooling rate (solid), from a calculation by Dickinson (1973), is superposed.

Through (8.68), (8.81) determines T_s under radiative equilibrium. Surface temperature increases with the optical depth of the atmosphere, a manifestation of the greenhouse effect. The surface temperature under radiative-convective equilibrium, although smaller, is likewise controlled by τ_s , for example, through (8.70) and (8.71).

The atmosphere's optical depth depends on its composition through radiatively active species. As described in Sec. 1.2.4, several have increased steadily during the last two centuries, inclusive of anthropogenic contributions. Presented in Fig. 8.30 is an estimate of the respective increase in greenhouse warming of the Earth's surface during the last two centuries, inferred from proxy records of composition (Sec. 1.6.2). Carbon dioxide is greatest, increasing downwelling LW radiation by $\sim 1.5 \text{ Wm}^{-2}$. Methane, nitrous oxide, halocarbons, and ozone are comparatively rare. Yet, together, they introduce almost as much additional warming as CO_2 . Collectively, these anthropogenic gases represent additional warming of the Earth's surface of about 3 Wm^{-2} . For reference, the additional warming introduced by a doubling of CO_2 is, from Figs. 8.30 and 1.14, seen to be $\sim 4 \text{ Wm}^{-2}$.

Offsetting the increased greenhouse warming is increased cooling of the Earth's surface by contemporaneous changes of aerosol. Entering the energy budget through albedo, those changes have likewise been inferred, somehow, from proxy records. Supported by surface changes, the inferred changes of aerosol represent an offset of about -1.5 Wm^{-2} .

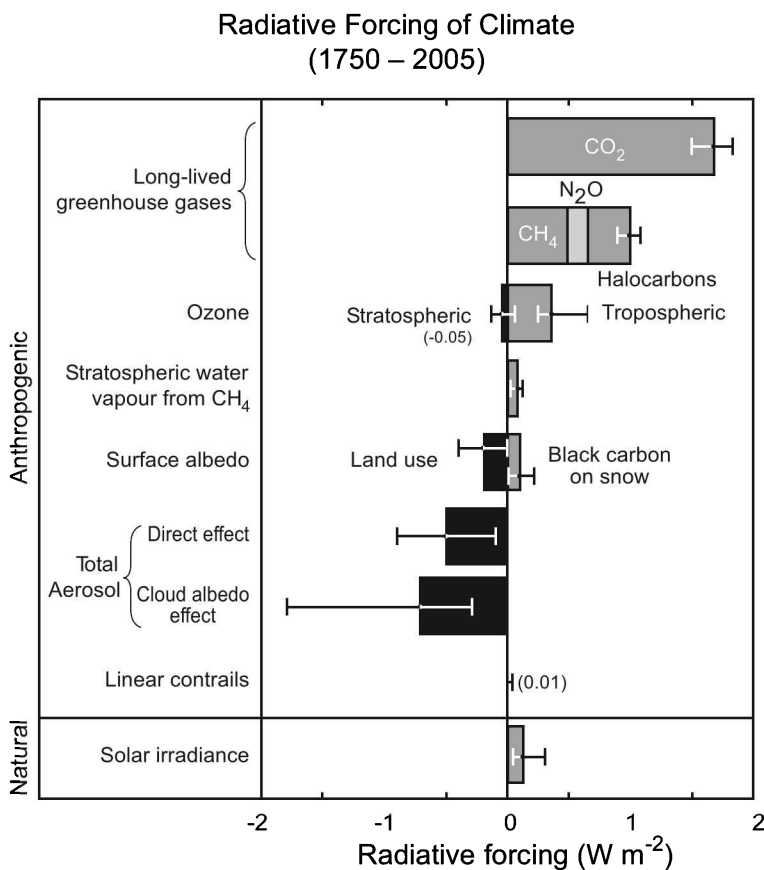


Figure 8.30 Inferred increase in radiative warming of the Earth's surface during the last two centuries from several influences. Not accounted for are changes of water vapor and cloud. Adapted from IPCC (2007).

The residual, $+1.5 W m^{-2}$, represents net warming. It is about 0.5% of the $327 W m^{-2}$ of overall downwelling LW radiation that warms the Earth's surface (Fig. 1.32). The vast majority of that warming is contributed by water vapor. Together with cloud, it accounts for 98% of the greenhouse effect. How water vapor has changed in relation to changes of the comparatively minor anthropogenic species (Fig. 8.30) is not known.

The additional surface warming introduced by anthropogenic increases in greenhouse gases amounts to about 75% of that which would be introduced by a doubling of CO₂. Arrhenius' estimate of 5–6°K for the accompanying increase of surface temperature (Sec. 1.2.4) then translates into $\sim 4^{\circ}K$. Yet, the observed change of global-mean temperature since the mid nineteenth century is only about $1^{\circ}K$ (Sec. 1.6.1). The discrepancy points to changes of the Earth-atmosphere system (notably, involving the major absorbers, water vapor and cloud) that develop in response to imposed perturbations, like anthropogenic emission of CO₂.

8.7.1 Feedback in the climate system

Surface temperature depends on the atmosphere's optical depth. The latter, in turn, depends on atmospheric composition through radiatively active species. Those

constituents are produced and destroyed by surface processes. Water vapor is produced at ocean surfaces through evaporation. Carbon dioxide is produced by decomposition of organic matter. Ozone is destroyed at the Earth's surface through oxidation. These and other processes that control radiatively active species are temperature dependent. This dependence introduces the possibility of feedback between the two sides of (8.81), which governs surface temperature through optical depth.

Because climate properties are interdependent, a perturbation of one will introduce changes in others. Secondary changes modify the direct response of surface temperature. The process through which this additional change is introduced is a *feedback mechanism*. If it reinforces the direct response of surface temperature, the process represents positive feedback. If it opposes the direct response of surface temperature, the process represents negative feedback. Climate feedback mechanisms are numerous. Some of the more important are highlighted.

Consider the energy balance at TOA. In the framework of Sec. 8.5, the net radiation (Sec. 1.4.2), averaged globally, is described by

$$\mathcal{R} = F_0 - F^\uparrow(\tau_s = 0), \quad (8.82.1)$$

where, with (1.30), (8.61) and (8.68),

$$F_0 = \frac{F_s}{4}(1 - \mathcal{A}) \quad (8.82.2)$$

$$F^\uparrow(0) = \epsilon\sigma T_s^4, \quad (8.82.3)$$

and ϵ represents an effective emissivity for the Earth-atmosphere system. Under equilibrium, $\mathcal{R} = 0$.

Suppose now that the system is perturbed, for example, by an increase of CO_2 that upsets the energy balance by $dF \text{ Wm}^{-2}$. This is equivalent to an increase in SW heating of dF : $F_0 \rightarrow F_0 + dF$. All components of the system will then undergo an adjustment to restore equilibrium. In terms of surface temperature, the change of net radiation is thus

$$d\mathcal{R} = 0 = \left(\frac{\partial \mathcal{R}}{\partial F} + \frac{\partial \mathcal{R}}{\partial T_s} \frac{dT_s}{dF} \right) dF, \quad (8.83)$$

where dF is the *direct radiative forcing*. As $\partial \mathcal{R} = \partial F$, (8.83) reduces to

$$\begin{aligned} \frac{dT_s}{dF} &= - \left(\frac{\partial \mathcal{R}}{\partial T_s} \right)^{-1} \\ &= \Lambda. \end{aligned} \quad (8.84)$$

The factor Λ defines the *climate sensitivity* with respect to a radiative perturbation dF . It measures the response of surface temperature to a perturbation dF in the energy balance: $dT_s = \Lambda dF$. With (8.82), the climate sensitivity becomes

$$\Lambda = \left(\frac{F_s}{4} \frac{\partial \mathcal{A}}{\partial T_s} + \frac{\partial F^\uparrow(0)}{\partial T_s} \right)^{-1}. \quad (8.85)$$

Albedo and upwelling LW flux in (8.85) depend on other properties, such as cloud cover and humidity. The contribution to Λ from the change of an individual property can be evaluated by holding others fixed.

Emission Feedback

An increase of T_s introduced through increased heating dF will produce an increase of LW emission at TOA. By cooling the Earth-atmosphere system, the increase of emission opposes the increase of T_s . It thus constitutes negative feedback. If albedo and effective emissivity in (8.82) are fixed, then emission is the only feedback present. The climate sensitivity then reduces to

$$\begin{aligned}\Lambda &= (4\epsilon\sigma T_s^3)^{-1} = (4\sigma T_e^3)^{-1} \\ &= \left[4 \frac{F^\uparrow(0)}{T_e}\right]^{-1} \\ &\cong 0.27^\circ\text{K} (\text{Wm}^{-2})^{-1}.\end{aligned}\tag{8.86}$$

A perturbation to the energy balance of 1 Wm^{-2} requires a response of surface temperature of 0.27°K . The former is equivalent to an increase in solar constant of $\sim 0.5\%$.

Temperature - Water Vapor Feedback

Owing to the strong temperature dependence in the Clausius-Clapeyron relation (4.38), saturation vapor pressure e_w increases sharply with T_s . An increase of surface temperature therefore increases the saturation vapor pressure e_w , which enables the vapor pressure e to increase. This increases τ_s and, from (8.81), further increases T_s . The response of water vapor thus reinforces the temperature increase that is introduced directly through dF . It constitutes positive feedback.

Such changes enter (8.82) through $F^\uparrow(0)$, implicitly through the effective emissivity ϵ . Representing them analytically is not feasible. Instead, the dependence on T_s of $F^\uparrow(0)$ has been evaluated empirically. Satellite observations indicate that $F^\uparrow(0)$ increases with T_s approximately linearly, $\frac{\partial F^\uparrow(0)}{\partial T_s}$ being in the range $1.55\text{--}1.85 \text{ Wm}^{-2}\text{K}^{-1}$ (North, 1975; Cess, 1976; Warren and Schneider, 1979). Supporting evidence comes from numerical simulations under radiative-convective equilibrium. Observations indicate that, despite large regional changes during its annual cycle, the relative humidity of the atmosphere as a whole varies little. Simulations with RH fixed reveal that $F^\uparrow(0)$ does not increase with T_s quartically, as in (8.82.3). Rather, it increases almost linearly, with $\frac{\partial F^\uparrow(0)}{\partial T_s} \cong 2.2 \text{ Wm}^{-2}\text{K}^{-1}$ (Manabe and Wetherald, 1967). If albedo remains fixed, then, inclusive of such feedback, (8.85) yields a climate sensitivity of

$$\Lambda \cong 0.55^\circ\text{K} (\text{Wm}^{-2})^{-1}.\tag{8.87}$$

As it rests on an observed relationship between $F^\uparrow(0)$ and T_s , (8.87) does not discriminate to water vapor; it also includes feedbacks from emission and cloud.⁵ Together, these feedbacks amplify the climate sensitivity. The response of surface temperature is now twice as large as would be required in the presence of only emission feedback (8.86).

A doubling of CO_2 over preindustrial levels would perturb the energy budget by 4 Wm^{-2} . From (8.87), it would thus be expected to produce warming of surface temperature of $\sim 2^\circ\text{K}$. The present increase of CO_2 over preindustrial levels corresponds

⁵ An increase of cloud acts in two opposing directions: It sharply increases optical depth, which increases T_s (8.81). Simultaneously, it sharply increases albedo, which decreases T_s (8.82.2). Although these influences enter at leading order, the complex dependence of cloud on temperature, humidity, and other factors leaves cloud feedback poorly understood.

to a LW perturbation of $\sim 1.7 \text{ Wm}^{-2}$ (Fig. 8.30). Via (8.87), it translates into an increase in surface temperature of $\sim 1 \text{ K}$. This value is broadly consistent with the observed increase of Global Mean Temperature over the last century (Sec. 1.6). It is not consistent with the increase of GMT during closing decades of the twentieth century. The latter implied a trend twice as great (Fig. 1.39). Indeed, after the turn of the century, GMT underwent a general decline for several years. Its increase during closing decades of the twentieth century is therefore unrepresentative of the secular variation of GMT. Because it pertains to a comparatively short interval, that warming is strongly contributed to by climate variability, as prevailed during earlier decades of the twentieth century.

Temperature - CO₂ feedback

Like other trace species, CO₂ is produced and destroyed at the Earth's surface. Involving a number of reservoirs and processes that are difficult to document, individual sources and sinks are poorly quantified. Net emission of CO₂, however, is determined unambiguously in the instrumental record (Fig. 1.43). As discussed in Sec. 1.6.2, the net emission rate of CO₂ varies coherently with T_s . An increase of T_s introduced by a radiative perturbation dF thus leads to an increase in the emission rate of CO₂, and, hence, cumulatively in CO₂. The latter, in turn, increases τ_s , reinforcing the increase of T_s introduced directly by dF . The dependence of CO₂ emission on temperature thus constitutes positive feedback. It is analogous to the positive feedback between temperature and water vapor. However, feedback between temperature and CO₂ is weaker - by two orders of magnitude (e.g., 4 Wm^{-2} vs. 327 Wm^{-2} ; cf. Fig. 1.32). Similar considerations apply to methane, which, like CO₂ is produced by decomposition of organic matter.

Such feedback can be treated in a manner analogous to water vapor. It is, in fact, already represented in the empirical relationship between T_s and $F^\uparrow(0)$. Derived from observed changes, that relationship accounts for feedback, not only with water vapor, but with all temperature-dependent species.

Increased temperature increases net emission of CO₂. Decreased temperature has the reverse effect. Each represents an imbalance between sources and sinks of CO₂. It is noteworthy that the positive sensitivity to temperature, $d\dot{r}_{CO_2}/dT_s$, is not restricted to small perturbations. As is evident in Fig. 1.43, the dependence on temperature applies to changes of \dot{r}_{CO_2} as large as 100%. Also noteworthy is that the correspondence applies to changes of temperature that are clearly of different origin. Following the eruption of Pinatubo, when SW heating decreased, \dot{r}_{CO_2} decreased by more than 50% (Fig. 1.27). During the 1997-1998 El Nino, when SST increased, \dot{r}_{CO_2} increased by more than 100%.⁶ To maintain stability, there must exist a negative feedback on CO₂, one that is sufficiently strong to bridle the enhancement of CO₂ emission by positive feedback from temperature. That negative feedback involves sinks of CO₂ at the Earth's surface (Sec 1.2.4). It is analogous to the negative feedback that bridles the enhancement of water vapor emission by positive feedback from temperature. Unlike CO₂, the negative feedback on water vapor involves sinks in the atmosphere.

The satellite record of GMT, in concert with the instrumental record of CO₂ is long enough to provide a population of climate perturbations, wherein the

⁶ The large increase of CO₂ emission then is consistent with the oceanic source of CO₂, which dominates transfers to the atmosphere (see Fig. 17.11)

Earth-atmosphere system was disturbed from equilibrium. Those perturbations reveal the sensitivity of CO₂ emission.

The climate sensitivity of CO₂ emission with respect to a change of temperature, $\frac{d\dot{r}_{CO_2}}{dT_s}$, is described in the same fashion as the climate sensitivity of temperature with respect to a change of radiative heating, $\frac{dT_s}{dF}$. Coherent changes of \dot{r}_{CO_2} and T_s in Fig. 1.43 imply a mean sensitivity

$$\begin{aligned} \frac{d\dot{r}_{CO_2}}{dT_s} &= c[\dot{r}_{CO_2}|T_s] \frac{\langle \dot{r}_{CO_2}^2 \rangle^{\frac{1}{2}}}{\langle T_s^2 \rangle^{\frac{1}{2}}} \\ &\approx 3.5 \frac{\text{ppmv/yr}}{\text{K}}, \end{aligned}$$

where $c[\dot{r}_{CO_2}|T_s]$ is the respective correlation and $\langle \rangle$ denotes time mean. Revealed by natural perturbations to the Earth-atmosphere system, the sensitivity accounts for much of the observed variation of CO₂ emission on interannual time scales (Fig. 1.43). It establishes that GMT cannot increase without simultaneously increasing CO₂ emission – from natural sources. Further, the coherent but out-of-phase changes in ¹²C and ¹³C (Fig. 1.43a, b) indicate that sources involved in interannual changes of CO₂ emission have isotopic composition similar to those involved in longer-term changes of CO₂ (Fig. 1.14). Their close correspondence to surface temperature begs the question: Does this dependence on T_s also contribute to changes of CO₂ emission on longer time scales?

During the satellite era, GMT in the record from MSU increased between 1979 and 2009 with a mean trend of $\frac{dT_s}{dt} \approx 0.125^\circ\text{K/decade}$. This represents warming over 3 decades of $\Delta T_s \approx 0.38^\circ\text{K}$. In concert with the above sensitivity, it implies an increase in CO₂ emission rate of $\Delta \dot{r}_{CO_2} \approx 1.3 \text{ ppmv/yr}$. The observed trend in \dot{r}_{CO_2} during the satellite era is $d\dot{r}_{CO_2}/dt \approx 0.26 \text{ ppmv/yr} \cdot \text{decade}$ (Fig. 1.43). It represents an effective increase in emission rate over 3 decades of $\Delta \dot{r}_{CO_2} \approx 0.8 \text{ ppmv/yr}$. Hence the increase in CO₂ emission rate anticipated from the increase in T_s alone is comparable to the observed increase in \dot{r}_{CO_2} . In fact, it exceeds the observed increase. Including increased emission from sources that are independent of temperature (e.g., anthropogenic sources) would magnify the anticipated increase even further.

During the twentieth century, GMT increased with a mean trend of $\sim 0.075 \text{ K/decade}$ (Sec. 1.6.1). With the above sensitivity, this translates into an increase in CO₂ emission rate of $\Delta \dot{r}_{CO_2} \approx 2.6 \text{ ppmv/yr}$. The observed increase in emission rate follows from Fig. 1.43, wherein \dot{r}_{CO_2} currently approaches 2.2 ppmv/yr, and from the ice core record (Fig. 1.14), which gives at the opening of the twentieth century $\dot{r}_{CO_2} \approx 0.2 \text{ ppmv/yr}$. Differencing yields an observed increase over the twentieth century of $\Delta \dot{r}_{CO_2} \approx 2.0 \text{ ppmv/yr}$. The increase in CO₂ emission anticipated from the increase in T_s alone is again comparable to the observed increase in \dot{r}_{CO_2} . If anything, it too exceeds the observed increase.⁷

The results for the two periods are in broad agreement. Together with the strong dependence of CO₂ emission on temperature (Fig. 1.43), they imply that a significant portion of the observed increase in \dot{r}_{CO_2} derives from a gradual increase in surface

⁷ The modest overestimate of the increase in emission rate (2.6 vs 2.0 ppmv/yr) is to be expected. On very-long time scales, mediating influences will eventually limit the response of natural sources to a change of temperature, invalidating the sensitivity that is apparent on interannual time scales. Yet, the comparatively-minor discrepancy with the observed increase in emission rate indicates that, even on the time scale of a century, the sensitivity observed on interannual time scales remains approximately valid.

temperature. Some of that warming is consistent with a rebound from the Little Ice Age (Moberg et al., 2005); cf. Fig. 1.45. Through integrated emission, the same implication must apply to CO_2 itself.

An estimate of this contribution follows from the temperature reconstruction in Fig. 1.45, which, as reconstructions go, is a conservative description of centennial changes during the last millennium. The rebound from the Little Ice Age involved systematic warming between the years 1600 and 1900 with a mean trend of $\sim 0.02^\circ\text{K}/\text{decade}$. This is about half of the mean trend over the entire instrumental record, which dates back to about 1850: $\sim 0.045^\circ\text{K}/\text{decade}$ (IPCC, 2007). With the above sensitivity of CO_2 emission, it implies an increase during the twentieth century of $\Delta r_{\text{CO}_2} \approx 55$ ppmv. The observed increase is ~ 80 ppmv. Such changes occur on time scales that are long compared to the length of instrumental records. Nevertheless, they are short compared to the time scales of heat transfer with the deep ocean (Chap. 17).

The preceding analysis follows from the observed sensitivity of CO_2 in a population of climate perturbations. It accounts for a significant component of observed changes during the satellite era and, more generally, during the twentieth century as a whole. The resemblance between observed changes of CO_2 and those anticipated from increased surface temperature also points to a major inconsistency between proxy records of previous climate. Proxy CO_2 from the ice core record (Fig. 1.13) indicates a sharp increase after the nineteenth century. At earlier times, proxy CO_2 becomes amorphous: Nearly homogeneous on time scales shorter than millennial, the ice core record implies virtually no change of atmospheric CO_2 . According to the above sensitivity, it therefore implies a global-mean climate that is “static,” largely devoid of changes in GMT and CO_2 . Proxy temperature (Fig. 1.45), on the other hand, exhibits centennial changes of GMT during the last millennium, as large as $0.5\text{--}1.0^\circ\text{K}$. In counterpart reconstructions, those changes are even greater (Section 1.6.2). It is noteworthy that, unlike proxy CO_2 from the ice core record, proxy temperature in Fig. 1.45 rests on a variety of independent properties. In light of the observed sensitivity, those centennial changes of GMT must be attended by significant changes of CO_2 during the last millennium. They reflect a global-mean climate that is “dynamic,” wherein GMT and CO_2 change on a wide range of time scales. The two proxies of previous climate are incompatible. They cannot both be correct.

Ice – Albedo Feedback

Coverage by ice is a significant contributor to the Earth’s albedo. It decreases with an increase of temperature. By (8.82.2), this increases F_0 . From (8.81), it further increases T_s . The response of ice cover thus reinforces the temperature increase that is introduced directly by dF . Like water vapor, it constitutes positive feedback.

We consider this influence in a simple model of the Earth-atmosphere system, one that includes interaction between radiative transfer and ice. Albedo is prescribed to vary with the idealized temperature dependence

$$\mathcal{A} = \begin{cases} \mathcal{A}_0 = \text{const} & T_s < 230 \text{ K} \\ \mathcal{A}_0 - \eta(T_s - 230) & 230 \leq T_s \leq 270 \text{ K} \\ \mathcal{A}_0 - \eta(270 - 230) = \text{const} & T_s > 270 \text{ K.} \end{cases} \quad (8.88)$$

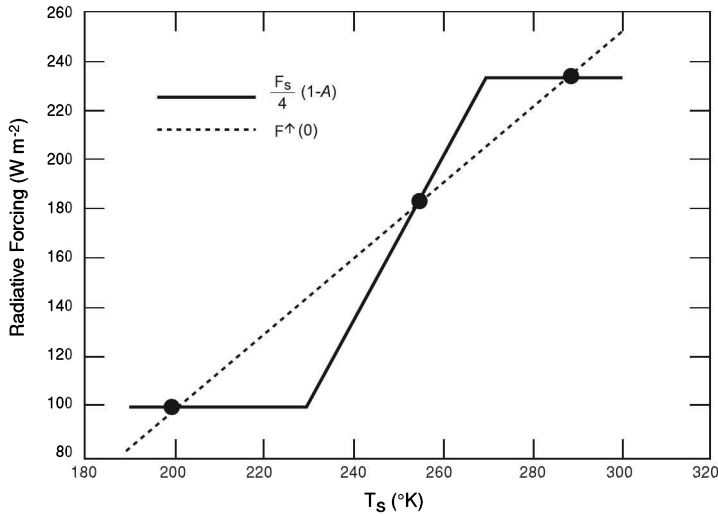


Figure 8.31 SW heating (solid) and LW cooling (dashed) at TOA, as functions of surface temperature T_s , in a simple climate system that includes ice-albedo feedback. With simplified dependence of albedo (8.88) and OLR (8.90), the values $\mathcal{A}_0 = 0.70$, $\eta = 0.01$, $F_0^\uparrow(0) = -212 \text{ Wm}^{-2}$, and $\gamma = 1.55 \text{ Wm}^{-2}/^\circ\text{K}$ make $\frac{F_s}{4}\eta > \gamma$ and $\Lambda < 0$. The system then possesses three equilibrium states (solid circles), at which SW heating equals LW cooling. Adapted from Kiehl (1992).

For T_s warmer than 270°K , the system is unglaciated; albedo then reduces to that of cloud and other fixed contributors. For T_s colder than 230°K , it is fully glaciated. Between those extreme states, ice cover decreases linearly with increasing T_s . Equilibrium requires $\mathcal{R} = 0$ or

$$\frac{F_s}{4}(1 - \mathcal{A}) = F^\uparrow(0). \quad (8.89)$$

Adopting for $F^\uparrow(0)$ the linear dependence on T_s

$$F^\uparrow(0) = F_0^\uparrow(0) + \gamma T_s, \quad (8.90)$$

together with (8.88) and (8.89), then defines a system of three simultaneous equations in the unknowns \mathcal{A} , $F^\uparrow(0)$, and T_s . With \mathcal{A}_0 , η , $F_0^\uparrow(0)$, and γ specified, solutions correspond to those values of T_s at which curves for the left- and right-hand sides of (8.89) intersect. In those states, SW heating equals LW cooling and the glaciated system is in equilibrium. The behavior is illustrated in Fig. 8.31, which plots the left- and right-hand sides of (8.89) for values that make $\frac{F_s}{4}\eta > \gamma$. The idealized climate system then possesses three equilibrium states. The warmest ($T_s = 288^\circ\text{K}$) represents a state of no ice. (\mathcal{A} is then invariant under small changes of T_s .) The coldest ($T_s = 202^\circ\text{K}$) represents a state that is fully glaciated. (\mathcal{A} is then likewise invariant under small changes of T_s .) Between those extreme states is the intermediate equilibrium state ($T_s = 254^\circ\text{K}$). It corresponds to partial glaciation. \mathcal{A} then decreases with small increases of T_s .

The climate system governed by (8.88)–(8.90) involves just a couple of variables. Even though highly simplified, it is capable of complex and wide-ranging behavior. The system's response to perturbation falls into three categories (Fig. 8.32), according to the sensitivities to increased T_s of increased SW heating (via reduced \mathcal{A}) and increased LW cooling (via increased $F^\uparrow(0)$):

- (1) $\frac{F_s}{4}\eta < \gamma$ (Fig. 8.32a): SW heating (solid) increases with increasing T_s *slower* than does LW cooling (dashed). The system then possesses only one equilibrium state, at which the curves for SW heating and LW cooling intersect. Under these conditions, (8.85) gives

$$\Lambda > 0.$$

Perturbing the TOA energy balance by $dF > 0$ is equivalent to introducing an infinitesimal upward displacement to the curve for SW heating: $\frac{F_s}{4}(1 - \mathcal{A}) \rightarrow \frac{F_s}{4}(1 - \mathcal{A}) + dF$ (dotted). The system evolves to its perturbed state along the curve for LW cooling. Upon reaching the displaced curve for perturbed heating, equilibrium is reestablished. As is apparent from Fig. 8.32a, this involves surface warming: $dT_s = \Lambda dF > 0$. The latter is required to offset the positive perturbation dF (heating) via an opposing perturbation $d[\frac{F_s}{4}(1 - \mathcal{A}) - F^\uparrow(0)]$, which, under these conditions, is negative (cooling).

- (2) $\frac{F_s}{4}\eta = \gamma$ (Fig. 8.32b): SW heating increases with increasing T_s *at the same rate* as does LW cooling. The system then possesses a continuum of equilibrium states, at which the curves for SW heating and LW cooling coincide. Under these conditions, (8.85) gives

$$\Lambda = \infty.$$

Perturbing the TOA energy balance by $dF > 0$ is again equivalent to an infinitesimal upward displacement of the curve for SW heating. However, because it is parallel to the curve for LW cooling, reestablishing equilibrium now requires a finite change of temperature. The system must evolve along the curve for LW cooling to a temperature warmer than the extremal temperature: $T_s > 270$ K, where it reaches the curve for perturbed heating and equilibrium is reestablished. Although stimulated by an infinitesimal perturbation in radiative heating, the temperature change $dT_s = \Lambda dF$ required to offset it and restore equilibrium is finite.

- (3) $\frac{F_s}{4}\eta > \gamma$ (Fig. 8.32c): SW heating increases with increasing T_s *faster* than does LW cooling. The system then possesses three equilibrium states, at which the curves for SW heating and LW cooling intersect (Fig. 8.31). Under these conditions, (8.85) gives

$$\Lambda < 0.$$

From the intermediate equilibrium state, the nearest perturbed equilibrium state is reached via surface cooling: $dT_s = \Lambda dF < 0$. The latter is required to offset the positive perturbation dF (heating) via an opposing perturbation $d[\frac{F_s}{4}(1 - \mathcal{A}) - F^\uparrow(0)]$, which, under these conditions, is negative (cooling) only if $dT_s < 0$.

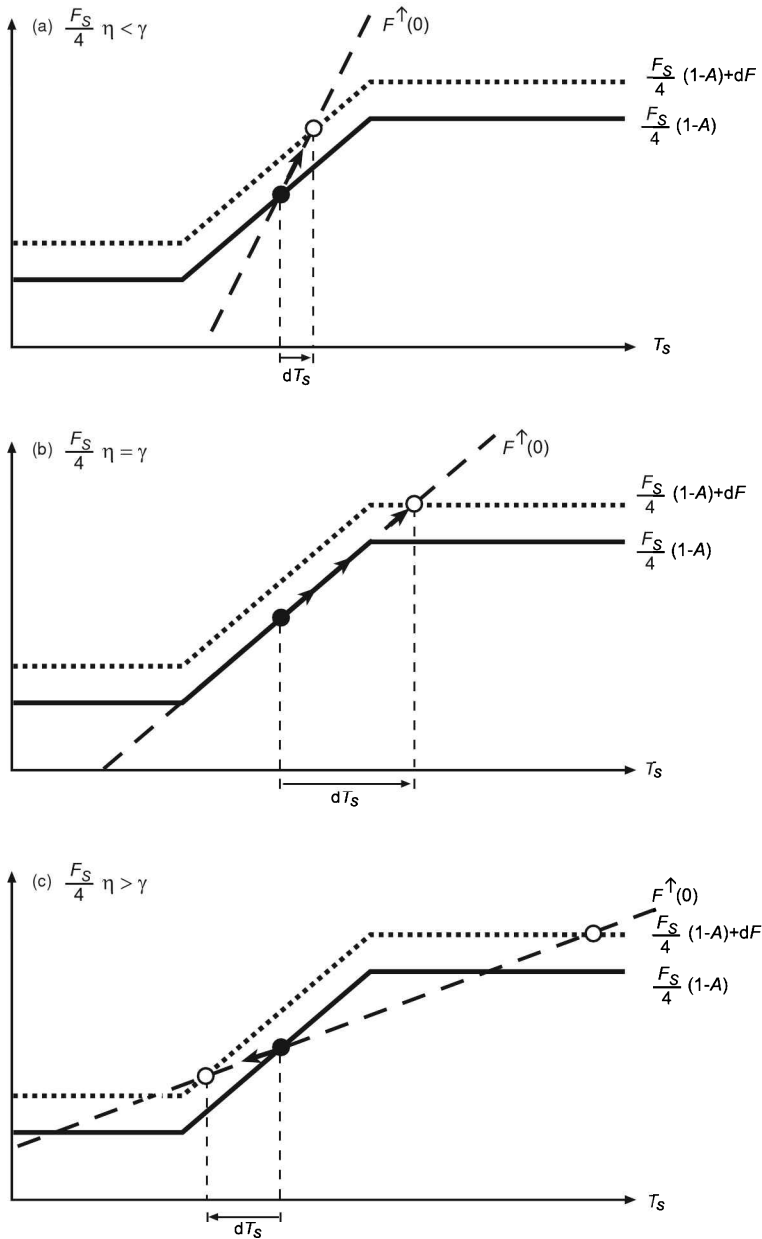


Figure 8.32 SW heating (solid) and LW cooling (dashed) at TOA as functions of surface temperature T_s in a simple climate system, inclusive of ice-albedo feedback, that is governed by (8.88)–(8.90). Superimposed are states of equilibrium, where SW heating equals LW cooling (solid circles) and the SW heating when perturbed infinitesimally by $dF > 0$ (dotted). (a) $\frac{F_S}{4} \eta < \gamma$: SW heating increases with increasing T_s *slower* than does LW cooling. When perturbed by infinitesimal heating, $dF > 0$, the system evolves to a new equilibrium state (open circle) through surface warming, $dT_s > 0$. (b) $\frac{F_S}{4} \eta = \gamma$: SW heating increases with increasing T_s *at the same rate* as does LW cooling. When perturbed by infinitesimal heating, $dF > 0$, the system evolves to a new equilibrium state through surface warming, $dT_s > 0$, that is finite. (c) $\frac{F_S}{4} \eta > \gamma$: SW heating increases with increasing T_s *faster* than does LW cooling. When perturbed by infinitesimal heating, $dF > 0$, the system evolves to the nearest new equilibrium state through surface cooling, $dT_s < 0$.

More sophisticated models of this form can possess one, two, or three equilibria, depending on the solar constant; see Hartmann (1994) for an overview. Under certain conditions, a partially glaciated state can be in equilibrium, but unstable. The system will then jump to another equilibrium state, one that is fully glaciated. Developing spontaneously (e.g., from an infinitesimal perturbation), this finite shift to the fully glaciated state occurs through positive feedback that goes unchecked (Sec. 8.7.2). The reverse change has been proposed in relation to methane that is sequestered beneath Arctic ice. Receding ice would release that methane, which would then reinforce greenhouse warming of the Earth's surface (Sec. 1.2.4). That, in turn, would accelerate the recession of ice, and so forth. How this hypothesis fits with the seasonality of CH_4 emission from Arctic tundra is unclear. During the freeze-in months of autumn, methane emission is as large as it is during unfrozen months of the entire summer (Mastepanov, 2008).

8.7.2 Unchecked feedback

Positive feedback such as that between temperature and water vapor reinforces small perturbations to the Earth-atmosphere system. Were it to continue without opposition, such feedback would produce a large shift of T_s , along with shifts of other properties that characterize climate. A paradigm of such changes is the so-called *runaway greenhouse effect*. It is used to explain the present state of the Venusian atmosphere.

The evolution of a planet's atmosphere occurs through slow discharge of gases from the planet's surface. In the Earth's atmosphere, water vapor operates in this manner. Atmospheric uptake of water vapor is limited by its saturation vapor pressure. The latter depends only on temperature (e.g., on surface temperature T_s). Because water vapor is responsible for atmospheric opacity, its vapor pressure e translates into optical depth τ_s . At saturation, $e = e_w(T_s)$. Plotting T_s against $\log(e_w)$ produces the saturation curve in Fig. 8.33 (dotted). It separates heterogeneous states of water from vapor phase alone.

Now the surface temperature of a planet is determined by the optical depth of its atmosphere τ_s and the incident solar flux F_0 , e.g., via (8.81) under radiative equilibrium or (8.70) under radiative-convective equilibrium. Thus the surface temperature T_s is a function of e , which is symbolic of τ_s . $T_s(e)$ defines a family of radiative-convective equilibrium curves, each one for a different value of F_0 . Two are plotted in Fig. 8.33. The curve for a particular F_0 can be thought to represent the evolution of a given planet. For small e , $T_s \cong (\frac{F_0}{\sigma})^{\frac{1}{4}}$. T_s then reflects the initial state of the planet, when all of its water resided at the surface in condensed phase (solid circle).

The situation for Earth is represented in the curve for Planet 1 (solid). Initially unsaturated, the atmosphere absorbs water vapor from the surface. This increases e , its LW opacity τ_s , and hence its surface temperature T_s . Continued absorption of water vapor and its positive feedback onto temperature then drive the state of the atmosphere to the right along the radiative-convective equilibrium curve. Eventually, the atmosphere's state encounters the saturation curve, where $e = e_w$ (open circle). Feedback is then checked by condensation, which returns H_2O to the planet's surface as fast as it is evaporated. Further increases of e are possible only through increases of $e_w(T_s)$. For the conditions of Planet 1, saturation is achieved at a surface temperature of about 287°K. These conditions describe a saturated system that contains multiple phases of water. They are characteristic of the present state of the Earth's atmosphere.

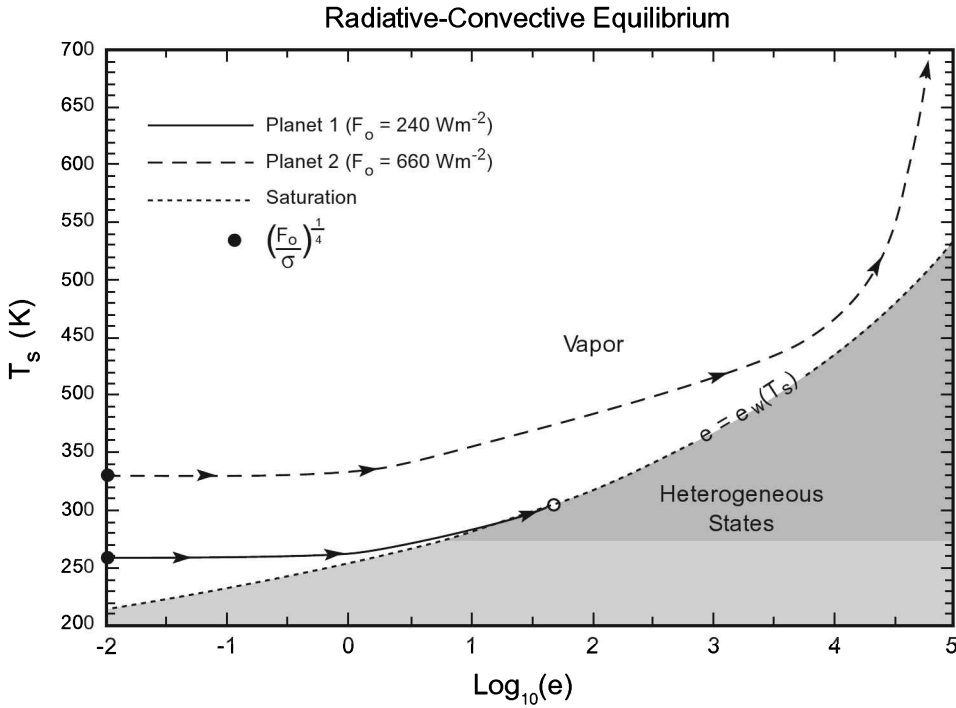


Figure 8.33 Surface temperature under radiative-convective equilibrium, as a function of vapor pressure for water (symbolic of atmospheric optical depth to LW radiation τ_s). Radiative-convective equilibrium curves shown for conditions representative of Earth (Planet 1) and Venus (Planet 2), which characterize atmospheric evolution from an initial state when all water resided at the surface in condensed phase. Positive feedback between temperature and water vapor drives the state of a planet’s atmosphere to greater temperature and humidity. For Planet 1 (solid), the radiative-equilibrium curve eventually encounters the saturation curve (dotted), where no more water is absorbed by the atmosphere and positive feedback ceases. However, the radiative equilibrium curve for Planet 2 (dashed) never encounters the saturation curve. It therefore continues to evolve through positive feedback until all water has been incorporated into the atmosphere. The calculation of radiative-convective equilibrium presumes a constant albedo and relates optical depth to vapor pressure as $\tau_s = 4 \frac{e(T)}{e(288\text{K})}$.

Saturation prevents H₂O from reaching great heights, where it would be photo-dissociated and ultimately destroyed when the resulting atomic hydrogen was lost to space (Sec. 1.2.2). Saturation of the atmosphere has thus maintained water on the Earth.

The curve for Planet 2 corresponds to Venus (dashed). It describes a very different evolution. Closer to the sun, Venus has a larger F_0 and hence a higher initial surface temperature. As before, the atmosphere’s state evolves to the right along the radiative-convective equilibrium curve. However, because of its higher initial temperature,

Planet 2 never encounters the saturation curve. Feedback of water vapor onto T_s then goes unchecked. Water evaporates from the planet's surface without bound. Eventually, all of the planet's H_2O has been absorbed by the atmosphere. Planet 2 has then achieved an optical depth and temperature that vastly exceed those present initially. This evolution is symbolic of Venus, which has mean surface temperature of $750^\circ K$ – some $400^\circ K$ warmer than would exist in the absence of its atmosphere.

Atmospheric composition implied by the evolution of Planet 2 also differs fundamentally from that of Planet 1. For initial water abundance comparable to that of Earth, the atmosphere of Planet 2 would achieve a state in which H_2O was its primary constituent. Without condensation, water vapor would then occupy all levels of the atmosphere. It would therefore suffer photodissociation by UV, allowing atomic hydrogen to escape to space. The free oxygen released could then recombine only at the surface through oxidation processes, which produce CO_2 . In this fashion, water would be systematically eliminated, replaced by carbon dioxide. This scenario is consistent with the present-day atmosphere of Venus, which is rich in CO_2 .

8.7.3 Simulation of climate

The feedbacks developed in Sec. 8.7.1 can amplify the response of the Earth-atmosphere system over its direct radiative response. In principle, those feedbacks are represented in global climate models. GCMs are, increasingly, coupled atmosphere-ocean models. The large heat capacity of the ocean, supported by transfers of latent heat, make climate simulations inherently sensitive to the ocean simulation (Sec. 17.2) Yet, limited observations of the deep-ocean circulation leave those simulations largely unvalidated.

GCMs predict that doubling CO_2 over pre-industrial levels will lead to warming of GMT by 3–7 K (IPCC, 2007). Those predictions are considerably greater than the warming anticipated by the observed feedback between temperature and water vapor (~ 2 K), which in turn is consistent with observed warming during the twentieth century. The magnified response represents an amplification of warming through feedback mechanisms inherent in GCMs. As climate projections rest on those internal feedbacks, which are poorly understood, the accuracy with which GCMs reproduce observed changes is pivotal.

Global climate models are sophisticated extensions of the idealized models considered above. Treatments of climate properties in different GCMs are as varied as they are complex. For some properties, like cloud cover, ice, and vegetation, they must resort to empirical relationships or simply ad hoc parameterization. For others, the governing equations cannot even be defined. Together with the ocean simulation, these limitations introduce errors, which can be substantial. Along with discrepancies between GCMs, they leave in question how faithfully climate feedbacks are represented (see, e.g., Tsushima and Manabe, 2001; Lindzen and Choi, 2009).

Discrepancies from observations and between models are smallest for global-mean properties, such as those represented in idealized models of climate (Sec. 8.7.1). Typified by GMT, global-mean properties benefit from a cancellation of error that is spatially incoherent: Fluctuations in one region operate independently of fluctuations in others. Regional properties, on the other hand, have error that is intrinsically greater, reflected in their greater variability (cf. Figs. 1.38, 1.39). For those, discrepancies with observations and between models can be substantial (IPCC, 2007).

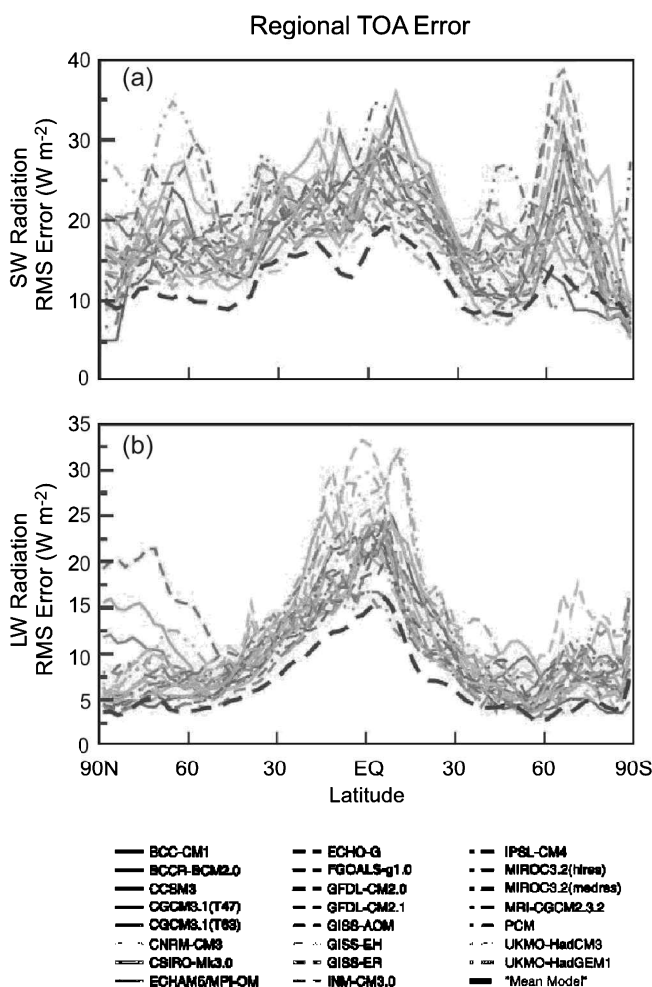


Figure 8.34 RMS error in time-mean regional fluxes at TOA simulated by several GCMs, relative to coincident fluxes observed by satellites in ERBE. (a) Error in reflected SW flux. (b) Error in outgoing LW flux. After IPCC (2007).

The accuracy of GCMs is reflected in the skill with which they simulate the TOA energy budget: the driver of climate. By construction, GCMs achieve global-mean energy balance. How faithfully the energy budget is represented locally, however, is another matter. The local energy budget forces regional climate, along with the gamut of weather phenomena that derive from it. This driver of regional conditions is determined internally - through the simulation of local heat flux, water vapor, and cloud. Symbolizing the local energy budget is net radiation (Fig. 1.34c), which represents the local imbalance between the SW and LW fluxes F_0 and $F^\uparrow(0)$ in the TOA energy budget (8.82). Local values of those fluxes have been measured around the Earth by the three satellites of ERBE. The observed fluxes, averaged over time, have then been compared against coincident fluxes from climate simulations, likewise averaged. Figure 8.34 plots, for several GCMs, the rms error in simulated fluxes, which have been

referenced against those observed by ERBE. Values represent the regional error in the (time-mean) TOA energy budget. The error in reflected SW flux, $\frac{F_r}{4} - F_0$ in the global mean (8.82), is of order 20 W m^{-2} (Fig. 8.34a). Such error prevails at most latitudes. Differences in error between models (an indication of intermodel discrepancies) are almost as large, $10\text{--}20 \text{ W m}^{-2}$. The picture is much the same for outgoing LW flux (Fig. 8.34b). For $F^\uparrow(0)$, the rms error is of order $10\text{--}15 \text{ W m}^{-2}$. It is larger for all models in the tropics, where the error exceeds 20 W m^{-2} .

The significance of these discrepancies depends on application. Overall fluxes at TOA are controlled by water vapor and cloud (Fig. 1.32) – the major absorbers that account for the preponderance of downwelling LW flux to the Earth's surface. Relative to those fluxes, the errors in Fig. 8.34 are manageable: Of order 10% for outgoing LW and 20% for reflected SW. Relative to minor absorbers, however, this is not the case. The entire contribution to the energy budget from CO_2 is about 4 W m^{-2} . Errors in Fig. 8.34 are an order of magnitude greater. Consequently, the simulated change introduced by increased CO_2 ($2\text{--}4 \text{ W m}^{-2}$), even inclusive of feedback, is overshadowed by error in the simulated change of major absorbers.

Discrepancies between GCMs arise from inaccuracies in climate properties and from differences in how those properties are represented. Much of the discrepancy surrounds the representation convection and its influence on water vapor and cloud, the absorbers that account for most of the downwelling LW flux to the Earth's surface. The involvement of convection is strongly suggested by models of radiative-convective equilibrium. Those simulations are inherently sensitive to how convection and cloud are prescribed. Cloud is especially significant to radiative considerations because it sharply modifies the atmosphere's scattering characteristics, which determine albedo, and its absorption characteristics, which determine optical depth.

SUGGESTED REFERENCES

- Fundamentals of Modern Physics* (1967) by Eisberg provides an excellent overview of Planck's theory and the emergence of modern physics.
- An Introduction to Atmospheric Radiation* (1980) by Liou includes discussions of solar variability and band models. An advanced treatment of radiative transfer is presented in *Atmospheric Radiation* (1995) by Goody and Yung, which includes a comparison of two-stream approximations. Applications of radiative transfer to remote sensing are discussed in *Atmospheric Physics* (2000) by Andrews.
- Aeronomy of the Middle Atmosphere* (1986) by Brasseur and Solomon and *Middle Atmosphere Dynamics* (1987) by Andrews et al. treat radiative transfer in the stratosphere and mesosphere.
- Global Physical Climatology* (1994) by Hartmann provides a detailed treatment of the energy budget, including climate feedbacks.
- Modeling the Atmospheric Component of the Climate System* (1992) by Kiehl contains a nice overview of the hierarchy of climate models. An overview of current models is presented in Bader et al. (2008).

PROBLEMS

1. Calculate the brightness temperature of the Earth based on (a) observed broadband OLR of 235 W m^{-2} , (b) the outgoing narrow-band flux (Fig. 8.5) at $\lambda = 11 \mu\text{m}$, (c) the outgoing narrow-band flux at $\lambda = 9.6 \mu\text{m}$, (d) the outgoing narrow-band flux at $\lambda = 15 \mu\text{m}$.

2. As the sun cools, its spectrum will shift toward longer wavelengths. Estimate the change in the Earth's equivalent blackbody temperature if the peak in the SW spectrum is displaced from its current position of $0.48 \mu\text{m}$ to a yellower wavelength of $0.55 \mu\text{m}$.
3. Orbital ellipticity brings the Earth some 3.5% nearer the sun during January than during July. This is responsible for the slight asymmetry of insolation apparent in Fig. 1.33. (a) Calculate the corresponding change in the Earth's equivalent blackbody temperature. (b) Discuss this change in relation to other changes that are introduced by hemispheric asymmetries of the Earth's surface and atmospheric circulation.
4. Consider a 250°K isothermal atmosphere that contains a single absorbing gas of uniform mixing ratio 0.05. The specific absorption cross section of that gas, as a function of wavelength, is given by

$$\sigma_{a\lambda} = 1.0e^{-\left(\frac{\lambda-\lambda_0}{\alpha}\right)^2} \text{m}^2\text{kg}^{-1},$$

where $\lambda_0 = 0.5 \mu\text{m}$ and $\alpha = .01 \mu\text{m}$. (a) Calculate, as a function of λ , the optical path length for downward-traveling radiation from the top of the atmosphere to the height z . (b) Plot the absorptivity of this layer, as a function of λ , for $z = 5, 3$, and 1 scale heights.

5. Show that a layer having optical thickness $\Delta u \ll 1$ has absorptivity $\Delta a = \Delta u$.
6. The top of the photosphere has a radius of $7 \cdot 10^8 \text{ m}$. If the mean Earth-sun distance is $1.5 \cdot 10^{11} \text{ m}$, calculate the equivalent blackbody temperature of the photosphere.
7. Venus has an albedo of 0.77 and a mean distance from the sun of $1.1 \cdot 10^{11} \text{ m}$. (a) Calculate the equivalent blackbody temperature of Venus. (b) Discuss this value in relation to the observed surface temperature of 750°K .
8. Consider an isothermal atmosphere that is optically thick to LW radiation. (a) How much of the LW radiance emitted to space at zero zenith angle originates in its uppermost 3 optical depths? (b) Determine the altitude range, in scale heights, occupied by the layer between optical depths of 0.1 and 3.0 under isothermal conditions.
9. A flat plate sensor on board a satellite behaves as a gray body with constant absorptivity a . Calculate its radiative-equilibrium temperature when the sensor faces the sun if (a) $a = 0.8$, (b) $a = 0.2$.
10. As for Prob. 8.9, but if the sensor has different absorptivities to SW and LW radiation of (a) a_{SW} and a_{LW} , respectively, (b) $a_{SW} = 0.2$ and $a_{LW} = 0.8$.
11. Under the conditions of Prob. 8.10, calculate the sensor's temperature as a function of time after the satellite has been rotated to face the sensor away from the sun. Presume that the sensor has an area of 0.1 m^2 and a specific heat of 2 J K^{-1} .
12. The hood of an automobile may be modeled as a gray flat plate with LW absorptivity 0.9 and the atmosphere as a gray layer with LW absorptivity 0.8. In this framework, estimate the hood's temperature for a solar zenith angle of 25° if (a) the automobile's color is light with a SW absorptivity of 0.2, (b) the automobile's color is dark with a SW absorptivity of 0.6. (c) What physical process would, in practice, mediate the values in (a) and (b)?
13. A simple model approximates the Earth's surface as a gray body and the atmosphere as a gray layer with LW absorptivities of 1.0 and 0.8, respectively. Use

this model to estimate the radiative-equilibrium temperature of the ground in the presence of 50% cloud cover (fully reflective) and under (a) perpetual summertime conditions over vegetated terrain: a solar zenith angle of 25° and the surface having a SW absorptivity of 0.6, (b) perpetual wintertime conditions over snow-covered terrain: a solar zenith angle of 60° and the surface having a SW absorptivity of 0.2.

14. An 80% solar eclipse occurs during morning, when the temperature would normally be 20°C . Use a gray atmosphere in radiative equilibrium with a black surface to estimate the change of surface temperature that would result were the eclipse to persist indefinitely (cf. Prob. 8.18).
15. Discuss how the competing influences of radiative transfer and convective mixing control stratification, in light of the diurnal variation of insolation.
16. The CO_2 absorption band is already saturated (Fig. 8.1). How then can increased levels of CO_2 produce global warming?
17. (a) Calculate the characteristic time scale of thermal damping in the troposphere based on $\rho_0 = 1.2 \text{ kg m}^{-3}$, the equivalent blackbody temperature $T_e = 255^\circ\text{K}$ (which is symbolic of where most OLR is emitted), and a characteristic depth of 5 km (which corresponds to the preceding temperature under radiative-convective equilibrium: Fig. 8.21). (b) What is the heat capacity of the surface under this approximation?
18. Use Newtonian cooling in Prob. 8.17 to estimate the amplitude ΔT_s of the diurnal cycle of surface temperature, where ΔT_s is half the nocturnal depression of temperature from its daytime maximum. Under equinoctial conditions and for a surface temperature maximum of 300°K , estimate ΔT_s for an optical depth (a) $\tau_s = 4$ that is representative of the global-mean conditions and (b) $\tau_s = 1$ that is representative of arid conditions.
19. Under radiative equilibrium, evaluate the change of solar constant that would increase the Earth's surface temperature by 1°K .
20. Use Figs. 1.13 and 1.24, in concert with Fig. 8.30, to determine an upper bound on total LW warming of the Earth's surface by CO_2 and CH_4 (the dominant greenhouse gases that have anthropogenic contributions). Apply the result, in concert with Fig. 1.32, to determine a lower bound on the respective LW warming by water vapor and cloud.
21. Consider a simple model of the Earth-atmosphere system, comprised of an isothermal gray layer with SW and LW absorptivities a_{SW} and a_{LW} , respectively, and a black underlying surface. Both are illuminated by a SW flux F_0 and are in thermal equilibrium. The energy budget may be formulated by tracing energy through repeated absorptions and reemissions by the surface and atmosphere. (a) Develop an arithmetic progression for the fractional energy absorbed by the atmosphere following successive transmissions from the surface to construct a series representation for the net energy absorbed by the atmosphere. (b) Use the series representation to calculate the radiative-equilibrium temperature of the atmosphere for $F_0 = (1 - \mathcal{A})\frac{F_s}{4} = 240 \text{ W m}^{-2}$ (Sec. 1.3), $a_{SW} = 0.20$, and $a_{LW} = 0.94$, which are representative of values in Fig. 1.32.
22. Use the results of Prob. 8.21 to determine the radiative-equilibrium temperature of the surface (a) for the conditions given, (b) in the absence of an atmosphere, (c) as in (a), but if the atmosphere's LW absorptivity is increased to $a_{LW} = 0.98$.

23. Estimate the level where the collisional half-width equals the Doppler half-width for (a) the water vapor absorption line at $\lambda^{-1} = 352 \text{ cm}^{-1}$ (Fig. 8.11), (b) the CO_2 absorption line at $\lambda^{-1} = 712 \text{ cm}^{-1}$.
24. Consider a discretely stratified atmosphere comprised of N isothermal layers, each of which is transparent to SW radiation but has an LW absorptivity a . Collectively, the layers are in radiative equilibrium with one another, an incident SW flux F_0 , and with a black underlying surface. (a) Derive expressions for the flux F_s emitted by the surface for $N = 1$, $N = 2$, and then generalize those expressions to arbitrary N . (b) Let $a = \frac{\tau_s}{N}$ and take the limit $N \rightarrow \infty$ to recover expression (8.81) for a continuously stratified atmosphere. (c) Verify that, even though they have different optical characteristics, the atmospheres for different N all have the same equivalent blackbody temperature. Explain why.
25. Derive expressions (8.38) for the upwelling and downwelling intensities in a plane parallel atmosphere as functions of optical depth.
26. Consider a gray atmosphere with a LW specific absorption coefficient k and an underlying surface with SW and LW absorptivities a_{SW} and a_{LW} , respectively. (a) Determine the distribution of radiative-equilibrium temperature as a function of optical depth, a_{SW} , and a_{LW} . (b) For $a_{LW} = 1$, discuss the limiting behavior: $a_{SW} \rightarrow 0$. (c) For $a_{SW} = 1$, discuss the limiting behavior: $a_{LW} \rightarrow 0$.
27. Show that the upwelling flux inside a gray atmosphere which is in radiative equilibrium with a black underlying surface is given by (8.71).
28. The Venusian atmosphere is composed chiefly of carbon dioxide, with $g = 8.8 \text{ m s}^{-2}$, $c_p = 8.44 \cdot 10^2 \text{ J kg}^{-1} \text{ K}^{-1}$ for CO_2 , a mean distance from the sun of 0.70 that for Earth, and an albedo of $\mathcal{A} = 0.77$. (a) Use radiative equilibrium and the observed surface temperature $T_s = 750^\circ\text{K}$ to estimate the optical depth of the Venusian atmosphere. (b) Calculate the temperature distribution under radiative-convective equilibrium based on (1) the profile of optical depth (8.69), with $h = \frac{RT_s}{g}$, (2) the presumption that the convective layer is warm enough to prevent saturation, and (3) in place of the constraint used in Sec. 8.5.2 to determine the tropopause height, requiring the temperature at the mid-level of the hydrostatically unstable layer to equal that under radiative equilibrium. (c) Plot the profiles of upwelling radiation, downwelling radiation, and radiative heating.

Aerosol and cloud

Radiative transfer is modified importantly by cloud. Owing to its high reflectivity in the visible, cloud shields the Earth-atmosphere system from solar radiation. It therefore introduces cooling in the SW energy budget of the Earth's surface, offsetting the greenhouse effect. Conversely, the strong absorptivity in the IR of water and ice sharply increases the optical depth of the atmosphere. Cloud thus introduces warming in the LW energy budget of the Earth's surface, reinforcing the greenhouse effect.

We develop cloud processes from a morphological description of atmospheric aerosol, without which cloud would not form. The microphysics controlling cloud formation is then examined. Macrophysical properties of cloud are developed in terms of environmental conditions that control the formation of particular cloud types. These fundamental considerations culminate in descriptions of radiative and chemical processes that involve cloud.

9.1 MORPHOLOGY OF ATMOSPHERIC AEROSOL

Small particulates suspended in air are produced and removed through a variety of processes. Those processes make the composition, size, and distribution of atmospheric aerosol widely variable (Table 9.1). Aerosol concentrations are smallest over ocean (Fig. 1.26), where a particle number density of $n = 10^3 \text{ cm}^{-3}$ is representative. They are greatest over industrialized continental areas, where $n > 10^5 \text{ cm}^{-3}$ is observed. These and other distinctions lead to two broad classes of tropospheric aerosol: *continental* and *marine*.

9.1.1 Continental aerosol

Continental aerosol includes (1) crustal species that are produced by erosion of the Earth's surface, (2) combustion and secondary components related to anthropogenic

9.1. Properties of atmospheric aerosol and cloud

Type of aerosol/cloud	Altitude (km)	Horizontal Scale (km)	Frequency of Occurrence	Composition	Mass Loading (mg/m ³)	Optical Depth (at 0.55 μm)	Mean Particle Radius (μm)	Principal Size Range (μm)
Stratocumulus, cumulus, nimbostratus cloud	1-18	10-1,000	0.5	Water, ice	1,000-10,000	~1-100	10-1,000	Variable
Altostratus cloud	7-16	10-1,000	0.3	Ice	10-100	~1	~10-100	Variable
Stratocumulus cloud	0-1	10-100	Sporadic	Water	10-100	1-10	~10	10-50
Stratospheric aerosol	0-10	1,000-10,000 (ubiquitous)	1	Sulfate, nitrate, minerals	0.01-0.1	~0.1	0.1-1	~0.3
Marine haze	0-1	100-1,000	0.3	Sea salt sulfate	0.1-1	0.1-1	0.5	~0.3
Storm clouds	0-3	10-1,000	Sporadic	Silicates, clays	< 1 - > 100	1-10	1-10	10-100
Stratocumulus cloud	5-35	100-10,000	Sporadic	Mineral ash, sulfates	< 1 - > 1,000	0.1-10	0.1-10	1-10
Stratocumulus cloud	0-10	1-100	Sporadic (from fires)	Soot, ash, tars	0.1-1	~0.1-10	0.1-1	~0.3
Stratospheric aerosol	10-30	1,000-10,000 (ubiquitous)	1	Sulfate	0.001-0.01	~0.01	0.1	~0.1
Stratospheric cloud	15-25	10-1,000	0.1 (winter only)	HNO ₃ /H ₂ O ice	0.001-0.01	~0.01-1	1-10	~1
Mesospheric aerosol	80-85	~200 (polar regions above 50°)	0.1 (summer only)	Ice	~0.0001	~0.0001-0.01	~0.05	~0.02
Desert dust	50-90	10-1,000	0.5-1	Minerals, carbon	~0.00001	~0.00001	≤ 0.01	Wide range incl. meteoric

WMO (1988).

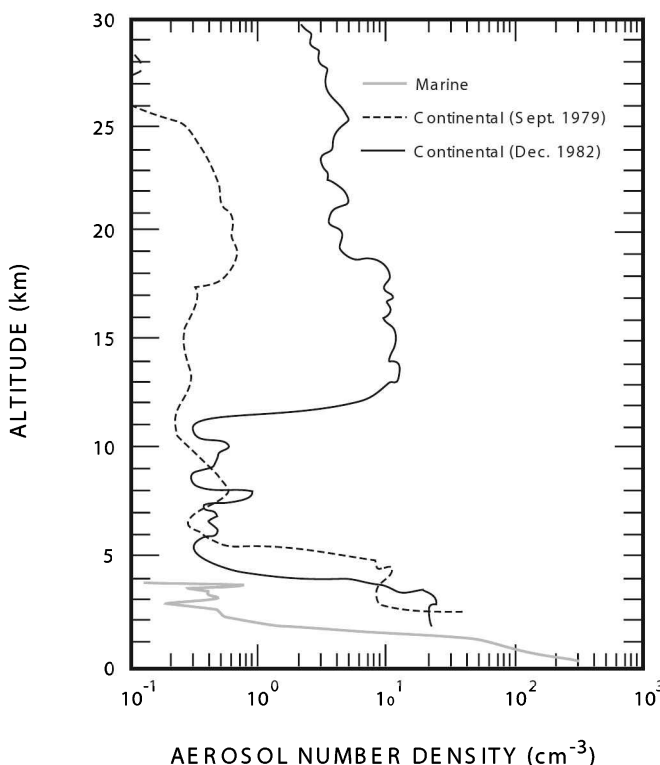


Figure 9.1 Aerosol number density n as a function of height in continental and marine environments. Profile of marine aerosol corresponds to particle radii $a > 0.5 \mu\text{m}$. The steep decrease of n above the marine boundary layer mirrors a similar decrease in dew point temperature (not shown). Profiles of continental aerosol correspond to $a > 0.15 \mu\text{m}$ and are shown before and after the eruption of El Chichon. Sources: Patterson (1982), WMO (1988).

activities, and (3) carbonaceous components consisting of hydrocarbons and elemental carbon. Crustal aerosol is produced in subtropical deserts like the Sahara, southwestern United States, and southeast Asia (see Fig. 9.43). Combustion and secondary aerosols originate chiefly in the industrialized regions of North America, Europe, and Asia. These continental source regions give the Northern Hemisphere greater number densities than the Southern Hemisphere. Carbonaceous aerosol has large sources in tropical regions due to agricultural burning and, through reaction with other species, to vegetative emissions.

Aerosol number density decreases upward from the surface (Fig. 9.1), which serves as a source for particles and gaseous precursors. In the troposphere, n has a characteristic height of only 2 km. This shallow depth reflects the absolute concentration of water vapor (Fig. 1.17) and the removal of aerosol by precipitation processes. In the stratosphere, n decreases upward slower.

Aerosol is characterized by its *size distribution*: $n(a)$, which denotes the number density of particles with radii smaller than a . Size distribution is a monotonically increasing function of a , with $n(a)$ approaching n as $a \rightarrow \infty$. The *size spectrum*: $\frac{dn}{da}$

AEROSOL SIZE SPECTRUM

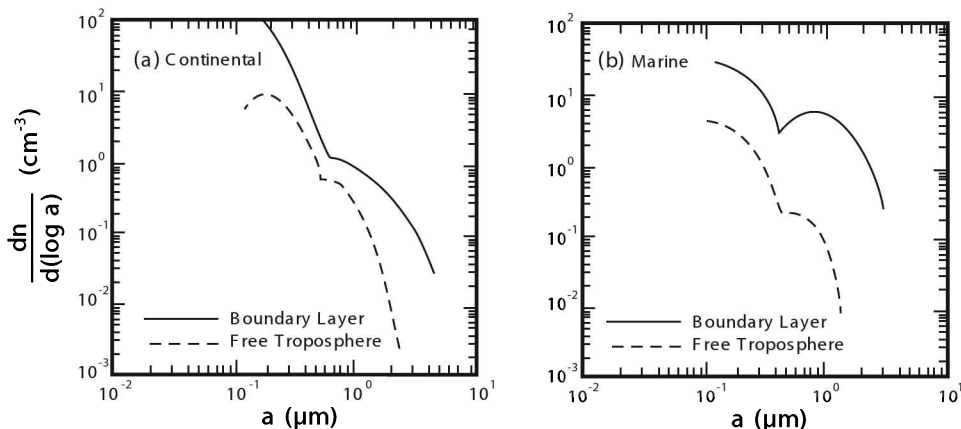


Figure 9.2 Aerosol size spectra, as functions of particle radius a , inside the planetary boundary layer and aloft for (a) continental environment and (b) marine environment. From Patterson (1982). Copyright by Spectrum Press (A. Deepak Publishing).

(Fig. 9.2) represents the contribution to $n = \int \frac{dn}{da} da$ from particles with radii between a and $a + da$. As for the spectrum of radiance (Fig. 8.1), plotting $\frac{dn}{d(\log a)}$ versus $\log a$ is area-preserving: Equal areas under the spectrum represent equal contributions to n from different size ranges. Crustal aerosol is comprised chiefly of windblown silicates. It involves particles with radii $a > 1 \mu\text{m}$. Secondary and combustion aerosols, which form through condensation, contain submicron-scale particles. These different production mechanisms give continental aerosol a bimodal size spectrum (Fig. 9.2a). Particles are concentrated in two distinct ranges of a . The size spectrum diminishes sharply with increasing a . The greatest contribution to n arrives from particles with $a < 0.1 \mu\text{m}$ - the so-called *Aitken nuclei*. Despite their relative abundance, Aitken nuclei constitute less than half of the surface area of aerosol (Fig. 1.26), which figures centrally in thermodynamic and chemical processes. They constitute only about 10% of its mass.

Combustion and secondary aerosols form through several mechanisms: (1) gas-to-particle conversion, wherein gaseous precursors undergo physical and chemical changes that result in particle formation, (2) direct condensation of combustion products, and (3) direct emission of liquid and solid particles. Gas-to-particle conversion is responsible for much of the sulfate and nitrate that dominate urban aerosol. That process is exemplified by a series of reactions that begins with the gaseous pollutant sulfur dioxide



followed by oxidation of HSO_3 to form H_2SO_4 vapor. Sulfuric acid vapor has a low saturation pressure. Therefore, it readily condenses, or *nucleates*, into liquid particles - especially in the presence of water vapor. Once formed, nuclei grow rapidly to radii of 0.01–0.1 μm . Larger aerosol can then develop through coagulation of individual nuclei

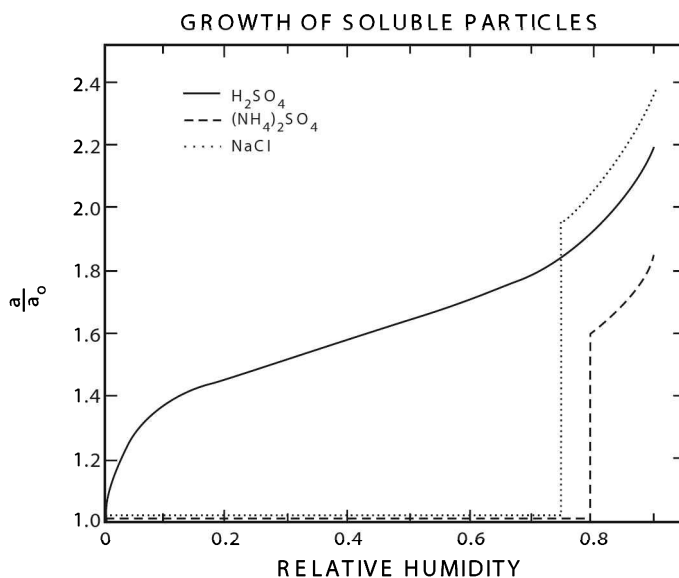


Figure 9.3 Aerosol particle radius as a function of relative humidity for liquid droplets and hygroscopic solids. Adapted from Patterson (1982). Copyright by Spectrum Press (A. Deepak Publishing).

to form *accumulation particles*. Sulfate aerosol is also produced through reaction of sulfur dioxide with ammonia inside cloud droplets. Upon evaporating, those droplets leave behind sulfate particles. In the troposphere, $(\text{NH}_4)_2\text{SO}_4$, NH_4HSO_4 , and H_2SO_4 are all present. In the stratosphere, supercooled H_2SO_4 droplets are the prevalent form of sulfate. Nitrogen emissions from combustion and natural sources lead to analogous products. Sulfate and nitrate are both important in the troposphere. However, nitrate is especially prevalent in urban areas because the prevalence of N_2 in air makes it an inevitable byproduct of combustion.

Carbonaceous aerosols such as soot consist of submicron-size particles. Natural sources include pollen and spores, as well as complex hydrocarbon vapors, like isoprene. Emitted via plant transpiration, those vapors react with oxides of nitrogen and subsequently nucleate.

Once formed, aerosol nuclei interact strongly with water vapor. Through condensation, liquid particles such as H_2SO_4 enlarge steadily with increasing relative humidity (Fig. 9.3). Conversely, hygroscopic solids such as NaCl and $(\text{NH}_4)_2\text{SO}_4$ remain dry below a threshold of $RH \cong 80\%$. Particles then dissolve, undergo a discontinuous enlargement, and thereafter enlarge with increasing RH similar to liquid aerosol.

Anthropogenic sources account for about 30% of aerosol production. The anthropogenic component of sulfate plays a role in cloud formation. Over urban areas, it exceeds 60% of production. The distribution of continental aerosol is strongly affected by sedimentation, especially for large crustal species. Fall speeds range from 50 cm s^{-1} for large silicates ($a \cong 50 \mu\text{m}$) to 0.03 cm s^{-1} for smaller ($a \cong 1 \mu\text{m}$) particles. As a result, large particles tend to be confined to a neighborhood of their source regions.

PRODUCTION OF MARINE AEROSOL

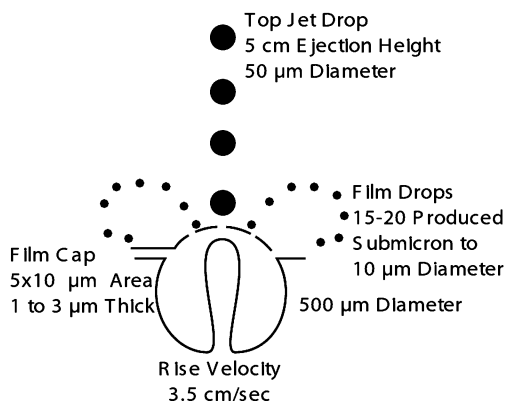


Figure 9.4 Formation of marine aerosol particles from a bursting ocean bubble. Adapted from Patterson (1982). Copyright by Spectrum Press (A. Deepak Publishing).

9.1.2 Marine aerosol

Composed primarily of sea salt, marine aerosol has a smaller overall concentration. Its number density decreases sharply above the boundary layer (Fig. 9.1), resembling the distribution of moisture. Like continental aerosol, marine aerosol has a size spectrum that is bimodal (Fig. 9.2b). It reflects two classes of droplets that form when ocean bubbles burst (Fig. 9.4). Water entrained into a bubble is ejected vertically in a stream of drops ($a \cong 25 \mu\text{m}$). Upon evaporating, they leave behind large ($a > 1 \mu\text{m}$) particles of sea salt. The thin film comprising the bubble's surface shatters, releasing droplets of $1 \mu\text{m}$ and smaller. Upon evaporating, they produce smaller particles of sea salt.

9.1.3 Stratospheric aerosol

Aerosol is introduced into the stratosphere through penetrative convection and volcanic eruptions. It also forms in situ through gas-to-particle conversion from precursors like SO_2 . That process maintains a background level of H_2SO_4 droplets, known as the *Junge Layer*. Its long residence time makes stratospheric aerosol fundamentally different from tropospheric aerosol. Small particulates with slow fall speeds are dynamically isolated by strong static stability. Sequestered from removal processes associated with precipitation, aerosol can survive in the stratosphere long after it is introduced. This feature enables stratospheric aerosol to alter SW absorption at the Earth's surface for extended durations. The great eruption of Krakatoa in 1883, which destroyed the Indonesian island of the same name, altered SW transmission and sunsets for 3 years afterwards (e.g., Humphreys, 1964). The eruption of Pinatubo had a similar influence, albeit weaker (Fig. 1.27).

Aerosol number density decreases upward in the stratosphere slower than in the troposphere (Fig. 9.1). Following major volcanic eruptions, aerosol number density actually increases into the stratosphere. This was evident following the eruption of El Chichon in 1982. A clear signature of El Chichon was registered in the transmission

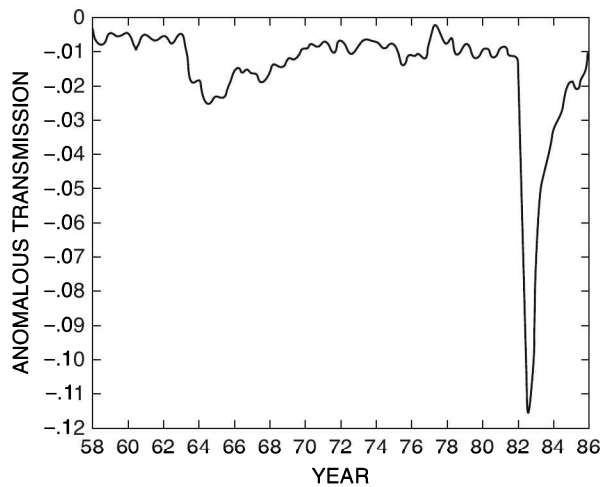


Figure 9.5 Change of direct solar transmission through the atmosphere, as observed over Hawaii by ground-based instruments. Note: Reduction of direct transmission is partially offset by an increase of diffuse transmission (e.g., Hoffman, 1988). The remainder is backscattered to space or absorbed. After WMO (1988).

of direct SW (Fig. 9.5), which decreased sharply in 1982. Through scattering and absorption, reduced transmission weakens SW heating of the Earth's surface. Such behavior was observed in temperature in the lower troposphere following Pinatubo (Fig. 1.27). Accompanying it was reduced water vapor in the upper troposphere, which is humidified by deep convection (Sec. 9.3.4). Increased extinction in volcanic debris can also increase radiative heating. It in turn can produce ascent of air in which the debris is suspended. Such was the case following the eruption of Pinatubo (WMO, 1995).

In the stratosphere, n has a characteristic vertical scale of 7 km. Resembling the scale height of air density, it suggests that relative concentration of aerosol (e.g., $\frac{n}{n_d}$) is approximately conserved for individual air parcels. Owing to the absence of removal processes, stratospheric aerosol behaves as a tracer. Figure 9.6 shows the cloud of volcanic debris introduced by the eruption of Mount St. Helens. Two months after the eruption, the cloud had dispersed zonally but not meridionally. Its confinement to high latitudes of the summer hemisphere follows from dynamical characteristics of the stratospheric circulation (Chap. 18).

9.2 MICROPHYSICS OF CLOUD

Large particles ($a > 1 \mu\text{m}$) are vulnerable to sedimentation. This process accounts for 10–20% of the overall removal of atmospheric aerosol. A more efficient removal mechanism is *washout*. Cloud droplets that form on aerosol particles carry them back to the Earth's surface. This removal mechanism is closely related to cloud formation.

9.2.1 Droplet growth by condensation

The simplest means of forming cloud is through *homogeneous nucleation*, wherein pure vapor condenses to form droplets. Suppose a small embryonic droplet forms

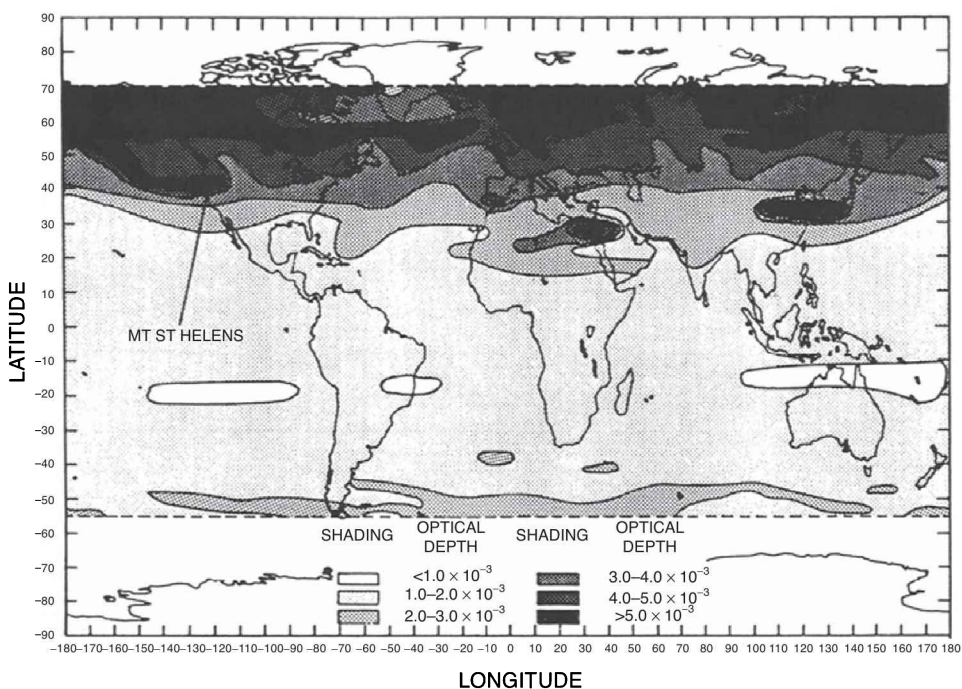


Figure 9.6 Mean aerosol optical depth at $1 \mu\text{m}$ 2 months after the eruption of Mt. St. Helens. Constructed by averaging satellite measurements from the Stratospheric Aerosol and Gas Experiment (SAGE) between July and August 1980. After Kent and McCormick (1984), copyright by the American Geophysical Union.

through chance collision of vapor molecules. The survival of that droplet is determined by a balance between condensation and evaporation. The equilibrium between the droplet and surrounding vapor is described by the Gibbs free energy (3.16.2), but with one modification. In addition to expansion work, a spherical droplet also performs work in association with its surface tension σ , which has dimensions of $\frac{\text{energy}}{\text{area}}$. Thus, σdA represents the work to form the incremental area dA of interface between vapor and liquid.

For a heterogeneous system that is comprised of the droplet and surrounding vapor, the fundamental relation for the Gibbs free energy (4.13) becomes

$$dG = -SdT + Vdp + (\mu_v - \mu_w)dm_v + \sigma dA, \quad (9.2.1)$$

where

$$\begin{aligned} dm_v &= -dm_w \\ &= -n_w dV_w, \end{aligned} \quad (9.2.2)$$

and n_w and V_w denote the number density and volume of the droplet.¹

¹ In phase transformations away from saturation, the system is out of chemical equilibrium. The chemical potentials in (9.2) therefore need not be equal.

The difference of chemical potential between the vapor and liquid phases may be expressed in terms of the vapor and saturation vapor pressures by appealing to the fundamental relation (3.16) for the individual phases.² Under an isothermal and reversible change of pressure de ,

$$d\mu_v = v_v de \quad (9.3.1)$$

for the vapor and

$$d\mu_w = v_w de \quad (9.3.2)$$

for the liquid. Subtracting gives

$$\begin{aligned} d(\mu_v - \mu_w) &= (v_v - v_w) de \\ &\cong v_v de. \end{aligned} \quad (9.4)$$

The gas law for an individual molecule of vapor is

$$ev_v = KT \quad (9.5)$$

(e.g., Lee, Sears, and Turcotte, 1973), where K is the Boltzmann constant. Introducing it into (9.4) gives

$$d(\mu_v - \mu_w) = KT dlne.$$

Integrating from the saturation pressure (at which $\mu_v = \mu_w$) then obtains

$$\mu_v - \mu_w = KT \ln\left(\frac{e}{e_w}\right). \quad (9.6)$$

Incorporating (9.6) into (9.2) and integrating from a reference state of pure vapor (with the vapor pressure maintained) yields the change of Gibbs free energy

$$\Delta G = -V_w n_w KT \ln\left(\frac{e}{e_w}\right) + \sigma A.$$

Thus, forming a spherical droplet of radius a corresponds to the change of Gibbs free energy

$$\Delta G = 4\pi a^2 \sigma - \frac{4}{3}\pi a^3 n_w KT \ln\left(\frac{e}{e_w}\right). \quad (9.7)$$

Figure 9.7 shows the free energy as a function of droplet radius for subsaturation ($\frac{e}{e_w} < 1$), saturation ($\frac{e}{e_w} = 1$), and supersaturation ($\frac{e}{e_w} > 1$). For subsaturation and saturation, G increases monotonically with droplet radius a . Now under isothermal isobaric conditions, a system approaches thermodynamic equilibrium by reducing its free energy (3.23). Consequently, droplet formation is not favored for $\frac{e}{e_w} \leq 1$. Those conditions characterize stable equilibrium and $G(a)$ an unbounded potential well. A droplet formed through the chance collision of vapor molecules will spontaneously evaporate, restoring the system to its original state. Under supersaturated conditions, $G(a)$ possesses a maximum at a critical radius a_c . Beyond that radius,

² Surface tension introduces a pressure difference between the droplet and surrounding vapor. It makes the pressure during a phase transformation variable. Strictly, that process should be treated as isothermal and isochoric, for which F is the appropriate free energy. However, discrepancies with G under isothermal and isobaric conditions are small enough to be ignored.

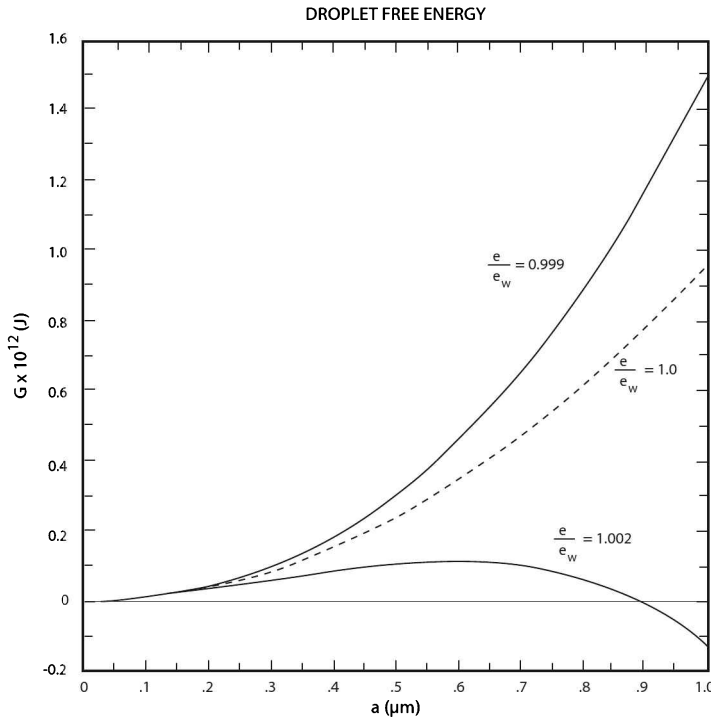


Figure 9.7 Gibbs free energy of a pure-water droplet produced through homogeneous nucleation, as a function of droplet radius. Shown for subsaturation ($\frac{e}{e_w} < 1$), saturation ($\frac{e}{e_w} = 1$), and supersaturation ($\frac{e}{e_w} > 1$).

free energy decreases. Condensation then becomes more efficient than evaporation. Under these circumstances, $G(a)$ characterizes conditional equilibrium and a bounded potential well. A droplet formed with $a < a_c$ evaporates back to the initial state ($G = 0$). But one formed with $a > a_c$ grows spontaneously through condensation of vapor. The system then diverges from its original state and the droplet is said to be *activated*.

The critical radius for a given temperature and vapor pressure follows by differentiating (9.7)

$$a_c = \frac{2\sigma}{n_w K T \ln\left(\frac{e}{e_w}\right)}. \tag{9.8}$$

Known as *Kelvin's formula*, (9.8) may be rearranged for the *critical* or *equilibrium supersaturation*

$$\epsilon_c = \left(\frac{e}{e_w}\right)_c - 1. \tag{9.9}$$

For supersaturation greater than ϵ_c , a droplet of radius a grows spontaneously through condensation. As shown in Fig. 9.8, small droplets have higher ϵ_c than large droplets. The larger is the droplet the smaller is the supersaturation necessary to maintain it. As $a \rightarrow \infty$, $\epsilon_c \rightarrow 0$. This behavior reflects the so-called *curvature effect* of surface tension

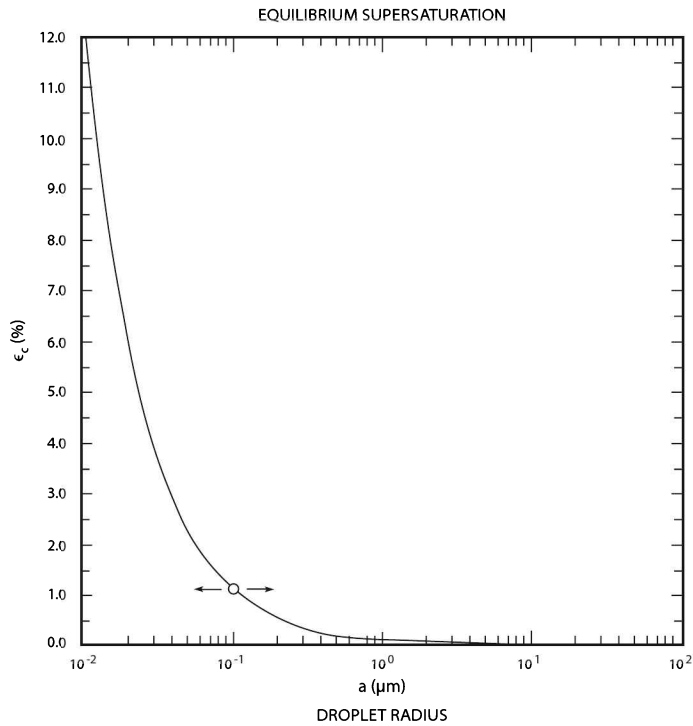


Figure 9.8 Critical supersaturation $\epsilon_c = \left(\frac{e}{e_w}\right)_c - 1$ at 278 K necessary to sustain a droplet of radius a . The critical value of ϵ describes an unstable equilibrium, wherein perturbations in a are reinforced. A perturbed droplet then either evaporates or grows through condensation away from the equilibrium state.

(9.2). It makes the equilibrium vapor pressure over a spherical droplet (a finite) greater than that over a plane surface ($a = \infty$).

The curve for ϵ_c describes an unstable equilibrium, which therefore cannot be maintained. Adding a vapor molecule drives a droplet of radius a above the equilibrium curve. The droplet then has a supersaturation ϵ above its equilibrium value ϵ_c . Therefore, the droplet continues to grow through condensation. Removing a water molecule has the reverse effect. The droplet then has a supersaturation below its equilibrium value ϵ_c . Therefore, the droplet continues to evaporate. In either case, a perturbed droplet evolves away from the equilibrium state.

Values of ϵ_c represent the barrier to droplet formation via the chance collection of vapor molecules. This is associated with the work required to overcome surface tension. Only droplets large enough for ϵ to exceed ϵ_c will survive. Even then, an embryonic droplet as large as $0.01 \mu\text{m}$ (large by molecular standards) still requires a supersaturation of 12% to survive. Yet, supersaturations exceeding 1% are rarely observed.³ Over

³ Near the tropical tropopause, where temperature can decrease below 190°K (Figs. 1.7; 7.8), supersaturation with respect to ice of several tens of percent have been reported (Spichtinger

most of the troposphere, $\epsilon = 0.1\%$ is typical. Consequently, the formation of most cloud cannot be explained by homogeneous nucleation.

Instead, cloud droplets form through *heterogeneous nucleation*, wherein water vapor condenses onto existing particles of atmospheric aerosol. Termed *cloud condensation nuclei* (CCN), such particles support condensation at supersaturations well below those required for homogeneous nucleation. Particles that are *wettable* allow water to spread over their surfaces. Such particles provide ideal sites for condensation because they then resemble a droplet of pure water, one which, at observed supersaturations, could not attain that size through homogeneous nucleation. According to Fig. 9.8, the larger such a nucleus is, the lower is its equilibrium supersaturation and hence the more it favors droplet growth through condensation. If other conditions are equal, large nuclei therefore activate sooner than small nuclei.

Even more effective are hygroscopic particles, such as sodium chloride and ammonium sulfate. In the presence of moisture, NaCl and $(\text{NH}_4)_2\text{SO}_4$ absorb vapor and readily dissolve (Fig. 9.3). The resulting solution has a saturation vapor pressure below that of pure water – because e_w is proportional to the absolute concentration of water molecules on the surface of the droplet. Consequently, a droplet containing dissolved salt favors condensation more than would a droplet of pure water of the same size. The so-called *Köhler curves* (Fig. 9.9) describe the equilibrium supersaturations for solutions that contain specified amounts of solute. A droplet that develops on a soluble nucleus therefore evolves along the curve corresponding to the fixed mass of that solute. The presence of NaCl sharply reduces the equilibrium supersaturation below that for a pure-water droplet (homogeneous nucleation), which is superimposed on Fig. 9.9 (dashed). For fixed a , ϵ_c decreases with increasing solute. Eventually, it becomes negative; that is, equilibrium is achieved for $RH < 100\%$. Nuclei that are more soluble activate sooner and are more favorable to droplet growth.

The equilibrium described by a given Köhler curve differs according to whether the droplet radius is larger or smaller than a at the maximum of the curve. The latter defines the threshold radius for growth, a_t . A droplet with $a < a_t$ is in stable equilibrium: Adding a vapor molecule drives the droplet along its Köhler curve to higher ϵ_c . The droplet's actual supersaturation ϵ then lies beneath its equilibrium value. The droplet therefore evaporates back to its original size. Removing a water molecule has the reverse effect. In either case, a perturbed droplet is restored to its original state. Droplets with $a < a_t$ cannot grow through condensation; they comprise *haze*. Conversely, a droplet with $a > a_t$ is in unstable equilibrium: Adding a vapor molecule drives that droplet along its Köhler curve to lower ϵ_c . The actual supersaturation then exceeds the droplet's equilibrium value. The droplet therefore continues to grow through condensation. Removing a water molecule has the reverse effect. In either case, a perturbed droplet evolves away from its initial state. The behavior is analogous to that of a pure-water droplet at its critical supersaturation (Fig. 9.8). Droplets with $a > a_t$ can enlarge into cloud drops.

Only a small fraction of tropospheric aerosol actually serves as CCN. Many particles are too small, not wettable, or insoluble. Continental aerosol is the dominant source of CCN because of its greater number density. It also has a high concentration of soluble sulfate, which is found in many cloud droplets. By contrast, sea salt is not common

et al., 2006; Jensen et al., 2004; Kramer et al., 2009). However, those values are unrepresentative of conditions that prevail elsewhere in the troposphere.

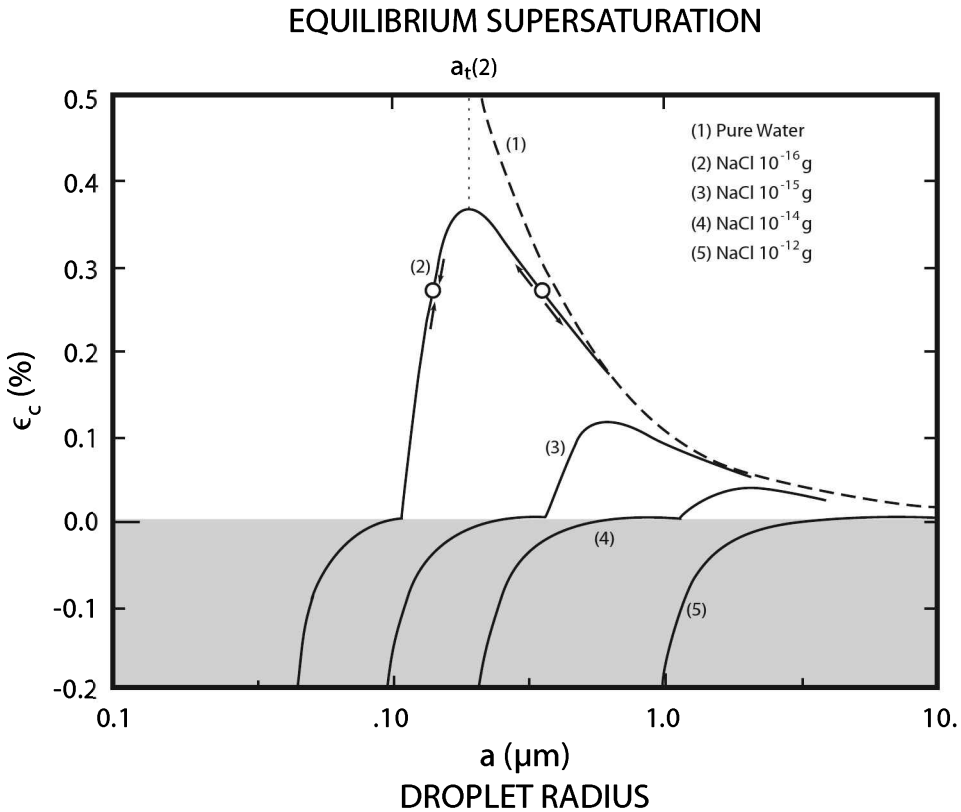


Figure 9.9 Critical supersaturation necessary to sustain a droplet of radius a and containing specified amounts of solute (solid). Critical supersaturation for a pure-water droplet (dashed) is superposed. For a given amount of solute, the corresponding *Köhler curve* describes different forms of equilibrium, depending on whether a is smaller or larger than the threshold radius at the maximum of the curve, a_t . For $a < a_t$, a droplet is in stable equilibrium: Perturbations in a are opposed by condensation and evaporation, which restore the droplet to its original state. For $a > a_t$, a droplet is in unstable equilibrium: Perturbations in a are reinforced, so the droplet either evaporates toward a_t or grows through condensation away from its original state.

in cloud droplets – not even over maritime regions. Instead, CCN in maritime cloud are composed chiefly of sulfate particles. Through anthropogenic emission, they have nearly doubled globally (IPCC, 1990). Sulfate aerosol also has a natural source: It forms through condensation of dimethylsulphide (DMS), which is emitted by phytoplankton at the surface of ocean.

Cumulus cloud has droplet number densities of order 10^2 cm^{-3} over maritime regions and 10^3 cm^{-3} over continental regions (Fig. 9.10). The greater number density of continental cloud mirrors the greater number density of continental aerosol. It also implies a smaller mean droplet size for continental cumulus. Nearly invariant between these cloud categories is the density of condensate, or *liquid water content* ρ_l (g m^{-3}). With droplets that are more numerous, continental cumulus therefore has a smaller

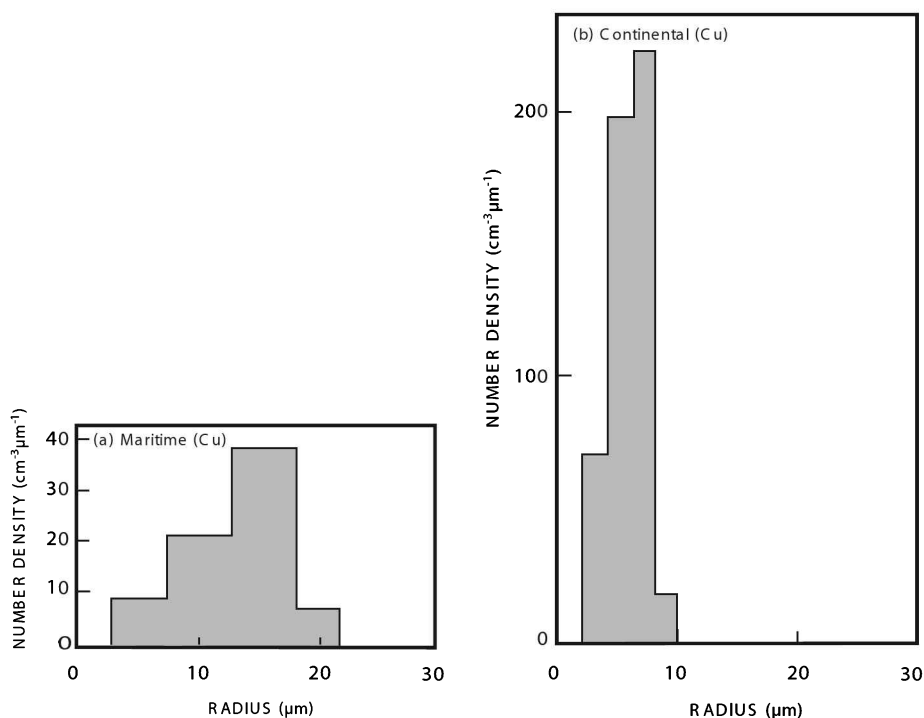


Figure 9.10 Droplet size spectra for cumulus cloud in (a) maritime environment and (b) continental environment. Note compressed scale in (b). Area under each curve reflects the overall number density n , which is an order of magnitude greater in continental cumulus than in maritime cumulus. *Source:* Pruppacher (1981).

mean droplet size. Droplets inside continental cumulus are also more homogeneous in size than droplets inside maritime cumulus: their size spectrum is narrow, concentrated about a preferred radius of $\sim 5 \mu\text{m}$. Maritime cloud droplets have a wider size spectrum. It reflects their formation on soluble sulfates, which activate over a wide range of sizes.

Liquid water content varies greatly across a cumulus complex (Fig. 9.11). Nevertheless, it remains strongly correlated with upward motion. ρ_l increases vertically to a maximum in the upper half of a cloud. Liquid water content then decreases sharply, vanishing at cloud top. It is noteworthy that ρ_l observed in nonprecipitating cloud seldom approaches adiabatic values ρ_l^{ad} , namely, values that would be attained by air parcels rising in isolation from environmental air. In fact, $\frac{\rho_l}{\rho_l^{ad}}$ rarely exceeds 0.5. At cloud top it vanishes. Implied is vigorous entrainment of drier environmental air at the walls and top of a cloud system (Sec. 7.4.2). For similar reasons, the greatest supersaturation is found in the core of a broad cumulus. Only there is updraft air isolated and hence undiluted by environmental air.

Before precipitation can occur, cloud droplets must enlarge by several orders of magnitude. After passing over the peak of its Köhler curve, a droplet can grow through condensation. The droplet's mass m then increases at the rate vapor diffuses across

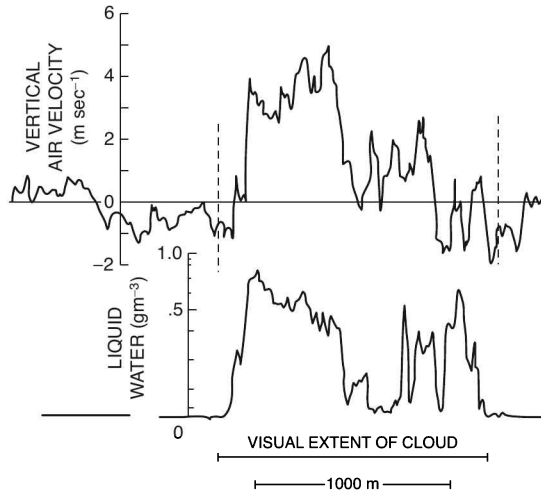


Figure 9.11 Variation of liquid water content ρ_l and vertical motion across a cumulus complex. After Warner (1969).

its surface

$$\frac{dm}{dt} = 4\pi r^2 v \frac{d\rho_v}{dr}, \quad (9.10)$$

where r denotes radial distance, v is the diffusion coefficient for water vapor, and ρ_v is its density. Integrating from the droplet radius $r = a$ to infinity and expressing the droplet's mass in terms of the density of liquid water ρ_w obtains

$$\frac{da}{dt} = \frac{v}{a\rho_w} [\rho_v(\infty) - \rho_v(a)].$$

With the gas law, this can be written

$$a \frac{da}{dt} = v \left[\frac{\rho_v(\infty)}{\rho_w} \right] \frac{e(\infty) - e(a)}{e(\infty)}. \quad (9.11)$$

For droplets larger than $1 \mu\text{m}$, the curvature and solute effects (Figs. 9.8, 9.9) are small enough to take $e(a) \cong e_w$. Then (9.11) may be expressed

$$a \frac{da}{dt} \cong v \left[\frac{\rho_v(\infty)}{\rho_w} \right] \epsilon. \quad (9.12.1)$$

If the right-hand-side is approximately constant, this gives

$$a(t) = \left\{ 2v \left[\frac{\rho_v(\infty)}{\rho_w} \right] \epsilon t \right\}^{\frac{1}{2}}. \quad (9.12.2)$$

Droplet growth through condensation (Fig. 9.12) is rapid at small radius, but slows as a droplet enlarges. This implies that a droplet population will become increasingly homogeneous in size as it matures. Although explaining the preliminary stage of droplet growth, condensation is too slow at large a to account for observed droplets in precipitating cloud.

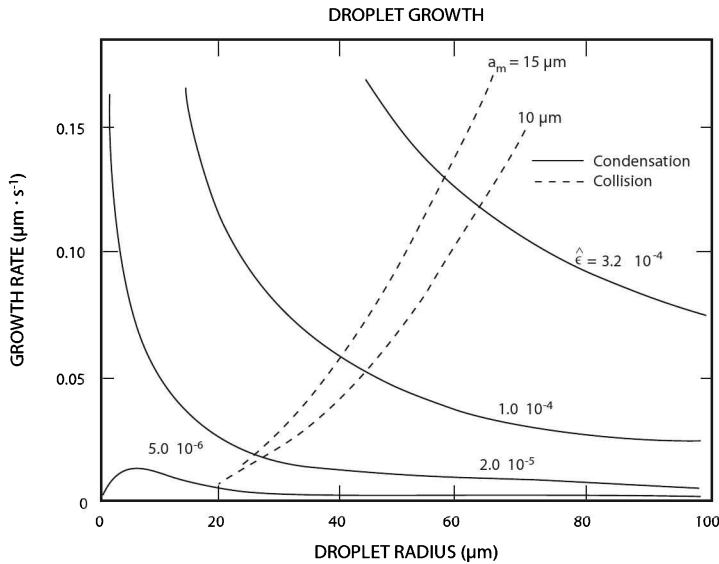


Figure 9.12 Droplet growth rate $\frac{da}{dt}$ due to condensation (solid) and collision (dashed) inside a cloud with liquid water content of $\rho_l = 1.0 \text{ g m}^{-3}$, as functions of droplet radius a . Growth rate due to condensation shown for different supersaturations $\hat{\epsilon} = \left[\frac{\rho_v(\infty)}{\rho_w} \right] \epsilon$ (see text). Growth rate due to collision shown for a droplet population whose size spectrum has a maximum at radius a_m . Adapted from Matveev (1967).

9.2.2 Droplet growth by collision

Cloud that lies entirely beneath the freezing level is called *warm cloud*. Droplets inside warm cloud also grow through coalescence, when they collide with other droplets. Large droplets have faster fall speeds than small droplets. A large *collector drop* then sweeps through a volume of smaller droplets. The collection efficiency E is a dimensionless quantity that reflects the effective collecting area posed to other droplets. It is small for collector radii less than $20 \mu\text{m}$. However, it increases sharply with collector size. For this reason, the wing of the droplet spectrum at large a is instrumental in forming precipitation.⁴

The rate a collector drop grows through collision with a homogeneous population of smaller droplets may be expressed

$$\frac{dm}{dt} = \pi a^2 E \rho_l \Delta w, \tag{9.13}$$

where a refers to the collector drop and Δw is the difference in fall speed between it and the population of smaller droplets. Rearranging as in the derivation of (9.12) yields

$$\frac{da}{dt} = \frac{E w \rho_l}{4 \rho_w}, \tag{9.14}$$

⁴ The collection efficiency also increases with the radius of the droplet population that is encountered by the collector drop (small droplets being swept around the collector drop without impact). Beyond a limiting droplet radius, however, coalescence is not favored.

Table 9.2. Fall speed of water drops

	Diameter (μm)	Fall Speed (cm/s)	Evaporation Distance (m)
Cloud Droplets	1	0.003	
	5	0.076	
	10	0.30	
	20	1.0	<1
	50	7.6	
Raindrops	100	27	
	200	72	150
	500	206	
	1000	403	
	2000	649	4200
	3000	806	
	5000	909	

Fall distance valid at 900 mb, 278 K, and $RH = 90\%$. Source: *Smithsonian Meteorological Tables* (1958), Smithsonian Institution, Washington DC; Battan (1984).

where $\Delta w \cong w$ has been presumed for the collector drop. Because w and E both increase with the radius of the collector drop, (9.14) predicts growth that accelerates with its size. For this reason, growth due to collision eventually dominates that due to condensation. According to Fig. 9.12, droplets smaller than $20 \mu\text{m}$ grow primarily through condensation (solid) – because fall speeds are small enough for such droplets to move in unison with the surrounding air, making collisions infrequent. However, droplets larger than $20 \mu\text{m}$ grow primarily through collision (dashed). Because it favors large drops, collision tends to broaden the size spectrum from the homogeneous population that is favored by condensation. It thus provides a mechanism for a small fraction of the drops, those of large radius, to grow much faster than the rest of the population.

Radii with $a > 100 \mu\text{m}$ dominate precipitation that reaches the ground. As seen in Table 9.2, they achieve fall speeds faster than 1 m/s. During their descent to warmer levels, cloud droplets and precipitation drops experience evaporation. The distance for complete evaporation increases with radius. Cloud droplets ($a < 100 \mu\text{m}$) are completely evaporated after falling less than a meter. Hence, they are too small to comprise precipitation. Rain drops ($a > 100 \mu\text{m}$) can fall several hundred meters before evaporating. Yet, even the largest fall only ~ 5 km before evaporating. For this reason, cloud with base higher than this value, prevalent over desert, produce no rainfall – irrespective of the column abundance of water. Marine cumulus have larger droplet sizes due to comparatively fewer CCN. Consequently, they are more likely to precipitate than continental cumulus of similar dimension. Collision also tends to limit the size of cloud droplets to a couple of millimeters because, during impact, larger drops fragment into smaller ones.

9.2.3 Growth of ice particles

Cloud that extends above the freezing level is called *cold cloud*. Cold cloud is seldom composed entirely of ice. In fact, nearly 50% of cloud warmer than -10°C and nearly

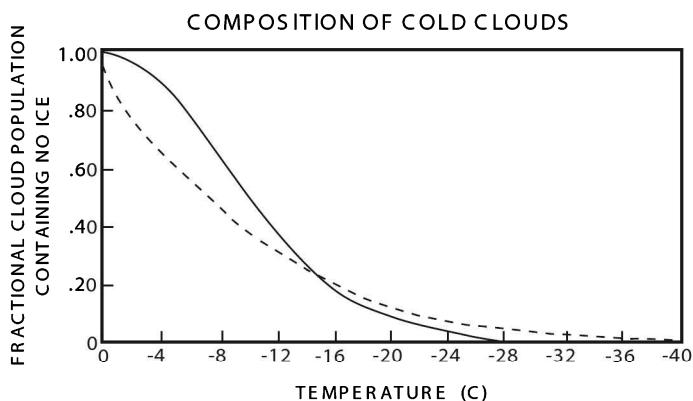


Figure 9.13 Fractional occurrence of unglaciated cloud, as a function of temperature, observed over Germany (solid) and Minnesota (dashed). *Source:* Pruppacher (1981).

all cloud warmer than -4°C contains no ice at all (Fig. 9.13). Droplets in such cloud remain in the metastable state of supercooled water (Prob. 3.2). However, 90% of cloud colder than -20°C contains some ice.

The reason why cold cloud is not entirely *glaciated* is that ice can form only under special circumstances. Homogeneous nucleation of ice involves the formation of an ice embryo through chance collection of water molecules inside a droplet. This process is not favored at temperatures above -36°C . At those temperatures, the free energy of formation increases for embryos smaller than a critical dimension - analogous to homogeneous nucleation of droplets (Fig. 9.7). The free energy then characterizes conditional equilibrium and a bounded potential well. An ice embryo smaller than the critical dimension for formation by chance collection of water molecules will then disperse. At temperatures above -36°C , the critical dimension for homogeneous nucleation is greater than $20\ \mu\text{m}$. Consequently, homogeneous nucleation of ice occurs only in very high cloud (see Fig. 7.7).

Heterogeneous nucleation occurs when water molecules collect onto an existing particle of a special type. Known as a *freezing nucleus*, it has molecular structure similar to ice. An ice embryo formed in this manner assumes the size of the freezing nucleus. It can therefore grow at temperatures much warmer than that permitting homogeneous nucleation. Maritime cloud, whose droplet size spectrum is broad with larger droplets (Fig. 9.10), glaciates more readily than does continental cloud. Crustal aerosol composed of clay is a prevalent form of freezing nucleus. But even at -20°C , fewer than one out of every hundred million aerosol particles serves as an ice nucleus. This leaves many cloud droplets supercooled. Aircraft encountering such cloud are therefore prone to icing.

Once formed, ice particles can grow through condensation of vapor, or *deposition*. The saturation vapor pressure with respect to ice is lower than that with respect to water (4.40). A cloud that is nearly saturated with respect to water may therefore be supersaturated with respect to ice.⁵ A particle's growth is therefore accelerated if it

⁵ Saturation mixing ratios with respect to water and ice typically differ by 10% or less. However, at very cold temperatures representative of the upper troposphere, the differences can be substantial (see Probs. 5.19, 18.4).


























	Nlc Elementary sheath		Pla Hexagonal plate		CP2b Bullet with dendrites
	N2b Combination of sheaths		Plb Crystal with sectorlike branches		R1a Rimed needle crystal
	Cla Pyramid		Plc Crystal with broad branches		R1b Rimed columnar crystal
	C1c Solid bullet		P1e Ordinary dendritic crystal		R1c Rimed plate or sector
	C1d Hollow bullet		P2a Stellar crystal with plates at ends		R2a Densely rimed plate or sector
	C1e Solid column		P2d Dendritic crystal with sectorlike ends		R4a Hexagonal graupel
	C1f Hollow column		P2g Plate with dendritic extensions		R4b Lump graupel
	C1g Solid thick plate		P4b Dendritic crystal with 12 branches		I1 Ice particle
			CP1a Column with plates		

Figure 9.14 Ice crystals observed in cold cloud. Adapted from Pruppacher and Klett (1978). Reprinted by permission of Kluwer Academic Publishers.

freezes. Similarly, in a mixed population, ice particles grow faster than droplets. Ice particles also grow through collision with supercooled droplets, a process known as *riming*. This produces a layered structure about the original ice particle. Beyond a certain dimension, the original structure is no longer discernible. The resulting particle is then referred to as *graupel*. Such particles assume irregular shapes, often proliferated with cavities. Lastly, ice particles can grow through coagulation, which is favored at temperatures above -5°C . These mechanisms lead to a wide array of shapes and sizes. Included are bullets, plates, slender needles, and complex configurations of hexagonal crystals (Fig. 9.14).

9.3 MACROSCOPIC CHARACTERISTICS OF CLOUD

9.3.1 Formation and classification of cloud

Most cloud develops through vertical motion, when moist air is lifted above its LCL and becomes supersaturated. Cloud falls into three broad categories:

- (1) *Stratiform (layered) cloud* develops from large-scale lifting of a stable layer. Characteristic of sloping convection, this process is exemplified by warm moist air that overrides cold heavier air along a warm front. Vertical motion that forms stratus cloud is of order 1 cm s^{-1} . The characteristic lifetime of stratus is of order a day.



Figure 9.15 Hook-shaped *cirrus uncinus*. After WMO (1969).

- (2) *Cumuliform* (piled) *cloud* develops from isolated updrafts. Associated with cellular convection, such cloud grows through positive buoyancy. The latter derives from sensible heat transfer from the Earth's surface and latent heat that is released during condensation. Both sources of buoyancy make cumulus cloud dynamic. Developing cumulus evolves on a time scale of minutes. Supporting updrafts are of order 1 m s^{-1} . However, in organized mature cells like *cumulus congestus* and *cumulonimbus*, they can be several tens of m s^{-1} . The characteristic lifetime of cumulus ranges from a few minutes to hours.
- (3) *Cirriform* (fibrous) *cloud* develops through either of the preceding forms of lifting. Found at great height, cirrus cloud is composed chiefly of ice particles. The larger particles descend in fall streaks or *mares' tails*. They give cirrus its characteristic fibrous appearance (Fig. 9.15).

Additional terms are used to specify particular cloud types. For example, *nimbo* and *alto* denote precipitation-bearing and mid-level, respectively. A complete tabulation is provided in WMO (1969).

Cumulus cloud develops in association with a buoyant updraft (Fig. 5.4). Known as a *thermal*, the updraft may be regarded as a succession of buoyant air parcels. Laboratory simulations reveal that the development of a thermal is closely related to its dissipation (Fig. 9.16a). A mass of buoyant fluid is transformed into a vortex

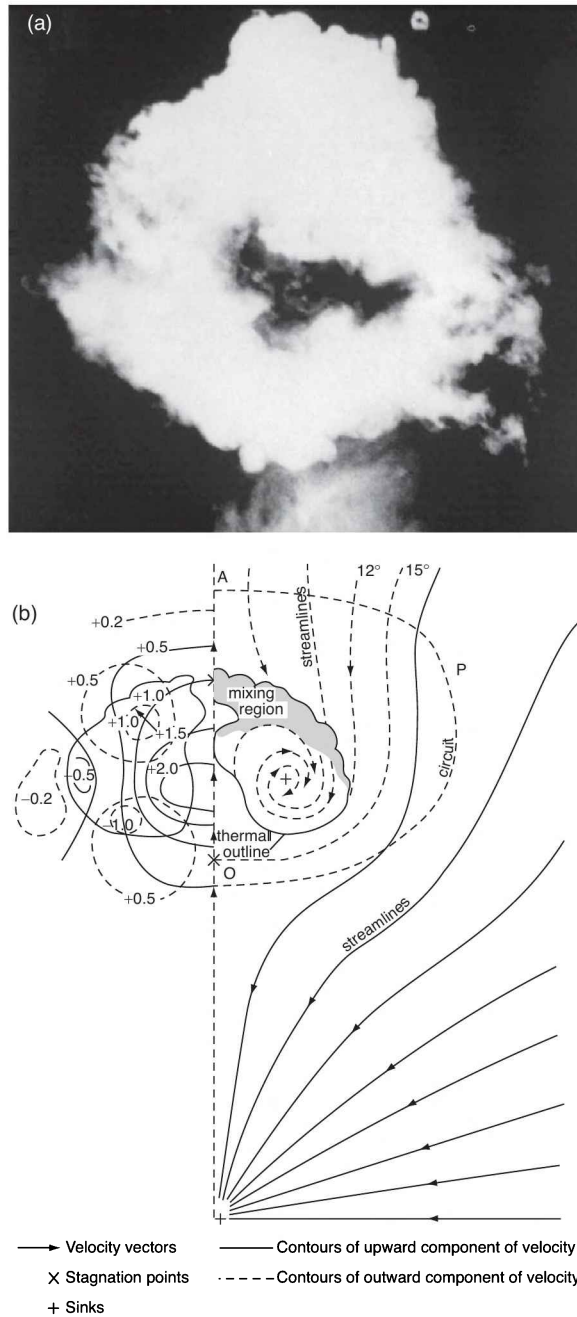


Figure 9.16 (a) Laboratory simulation of buoyant fluid inside a thermal. After advancing 1–2 diameters, a mass of buoyant fluid has been turned inside out. Environmental fluid entrained into its core produces a hole in the center of the buoyant mass. (b) Motion and streamlines about the moving body of fluid. Mixing with surrounding fluid causes the advancing mass to expand at an angle from the vertical of about 15° . Adapted from Scorer (1978).

ring (analogous to a smoke ring). The latter is characterized by a hole in its center. Punctuating its advancing surface is a complex array of protrusions or *turrets*. A mass of buoyant fluid penetrates into the surroundings only by entraining environmental fluid into its center (Fig. 9.16b). This process dilutes the buoyancy of a thermal. It also causes it to expand – at an angle of about 15° from the vertical. In the process, the buoyant mass is turned completely inside out after advancing only about $1\frac{1}{2}$ diameters. At that stage, the buoyant mass has become thoroughly mixed with surrounding fluid. Its buoyancy has then been destroyed. The direct involvement of vorticity is illustrated in the trajectory of individual fluid parcels. Some complete several circuits about the advancing vortex ring before falling out of its influence.

Expansion is less evident in cumulus cloud because the latter undergoes dissolution. Mixing with drier environmental air along the cloud's periphery leads to its evaporation. As observed, liquid water content is well below adiabatic values, mixing occurs throughout a cumulus cloud. Except for very broad cloud, some environmental air is communicated even into the core of cumulus. Consequently, a cumulus can grow to appreciable height only by expanding horizontally – enough to isolate saturated air in its core from drier environmental air that is mixed across its walls. Rising motion inside the updraft is compensated by gradual sinking motion in its environment. That subsidence is distributed over a broad area. It is promoted by evaporative cooling when saturated updraft air mixes with unsaturated environmental air and dissolves. By inhibiting ascent, subsiding motion in the environment organizes new parcels into the existing updraft, a process that favors cloud growth. Likewise, buoyant parcels are most likely to produce cloud growth if the region into which they advance has already been humidified and its stability reduced by preceding parcels.

A developing cumulus is marked by sharply defined turrets along its boundary, especially at its top. The turrets dissolve through evaporation as they penetrate into unsaturated air, replaced by new ones. The base of a developing cumulus is also sharply defined. It is level or even concave down – because warmer temperature in the thermal's core elevates the LCL (5.24). Conversely, a decaying cumulus is diffuse, rendered amorphous by turbulent mixing. That process leads to evaporation, quickly blurring sharp features that were created during the cloud's development. The cloud base may then become concave up, reflecting a reversal of vertical motion and the horizontal temperature gradient. Cloud matter that descends beneath the cloud base is termed *virga*. It is referred to as precipitation only if it reaches the ground (corresponding to drop radii $a > 100 \mu\text{m}$).

Thermal growth is inhibited by strong vertical stability. Figure 9.17 illustrates a thermal that encounters and partially penetrates through a layer of strong stability. After penetrating the stable layer, buoyant fluid continues to advance, eventually diffusing via detrainment and turbulent mixing with the surroundings. Fluid deflected by the stable layer fans out, forming anvil. Such behavior appears occasionally at intermediate levels in the troposphere (Fig. 9.18). It forms invariably when a cumulonimbus tower encounters stable air at the tropopause. Penetrative overshoots at the tropopause are accompanied by turbulent entrainment (Fig. 7.7). That process mixes tropospheric and stratospheric air, which eventually splash back to the LNB to occupy the resulting anvil.

The stratiform anvil of a mature cumulonimbus occasionally evolves into small cellular convection. LW emission at its top and absorption at its base destabilize the anvil layer, introducing convective overturning (Sec. 9.4.2). Domes of sinking virga then define *mammatus* (breast-shaped) *cloud* (Fig. 9.19). Their descent is promoted

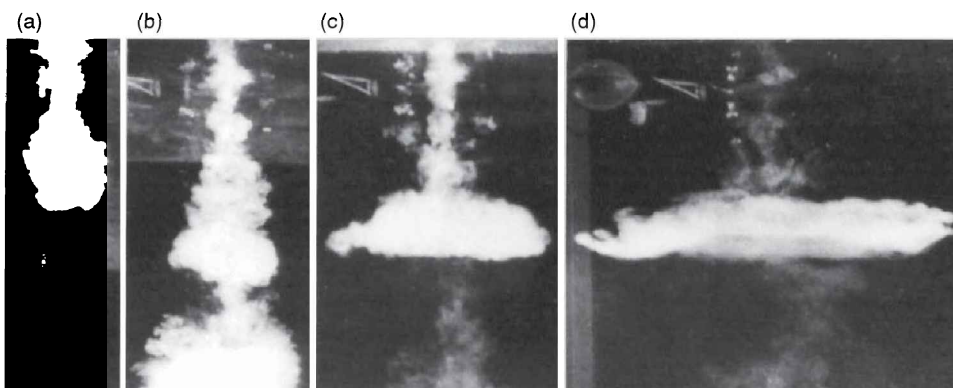


Figure 9.17 Laboratory simulation of a thermal encountering a layer of strong static stability. Upon reaching the stable layer, the negatively buoyant plume is detrained horizontally into an anvil. Fluid penetrating the stable layer continues to advance and eventually diffuses through turbulent mixing. After Scorer (1978).

by evaporative cooling. Spaced quasi-regularly, mammatus features resemble cells in Rayleigh-Bénard convection (Fig. 9.20), structure that develops in a shallow fluid that is heated differentially at its top and bottom. The horizontal dimension of Rayleigh-Bénard cells reflects the depth of the layer that is convectively overturned.⁶

Behaving similarly is stratiform cloud that is destabilized by absorption of LW radiation from the surface, producing radiative warming of its base, and emission to space, producing radiative cooling of its top. Instability introduced by such radiative heating and cooling can transform a continuous stratiform layer into an array of cumulus cells with a preferred horizontal dimension. *Stratocumulus* and *fair weather cumulus* that are found behind a cold front develop in this fashion (e.g., Fig. 9.21). Cold air that is advected over warmer surface is then heated at its base and cooled radiatively at its top.

Cloud also forms in the absence of buoyancy through forced lifting over elevated terrain. *Orographic cloud*, unlike other forms, does not mark a particular body of air. Rather, it marks a control volume through which air flows. *Lenticular* (lens-shaped) *cloud* forms inside an organized wave pattern (Fig. 9.22a). This smooth cloud, often forming in a stack of lenses, develops when the air stream is deflected over a mountain range. Each lens in Fig. 9.22a continually forms along its leading edge, where air is displaced above its LCL. It then dissolves along its trailing edge, where air returns below its LCL. Also known as *mountain wave cloud*, lenticular is often found at great height. On occasion, it is observed even in the stratosphere, where it is termed *nacreous cloud*. Vertical motion that accompanies wave cloud has been harnessed by sailplanes to achieve record-breaking altitudes – as high as 50,000'. The vertical structure of lenticular cloud is controlled by environmental conditions. When moisture is stratified

⁶ According to Rayleigh-Bénard theory, conversion of potential energy into kinetic energy becomes increasingly efficient with decreasing horizontal dimension of the cells. Energy conversion is eventually offset by frictional dissipation (e.g., mixing of momentum across the walls of the cells), which becomes important at horizontal dimensions comparable to the depth of the fluid. It is no accident that cumulonimbus cells have horizontal dimensions comparable to the depth of the troposphere.



Figure 9.18 A developing cumulus that encounters and penetrates through a shallow layer of strong static stability. Courtesy of Ronald L. Holle (Holle Photography).

in layers, lenticular cloud forms in a stack (Fig. 9.22b). Such vertical structure can result from towers of anomalous moisture that have been left upstream by prior convection. When sheared by the circulation, they fold to form layered structure.

The smooth form of lenticular characterizes air flow that is turbulence-free or *laminar*. That steady motion is often replaced just downstream by severe turbulence. It is



Figure 9.19 *Mammatus* cloud that develops from a cumulonimbus anvil. Supplied by the National Center for Atmospheric Research, University Corporation for Atmospheric Research, National Science Foundation.

Figure 9.20 Laboratory simulation of Rayleigh-Bénard convection. A fluid layer heated at its base evolves into a regular array of hexagonal cells with a preferred horizontal scale. Fluid ascends in the center of each cell and descends along its edges. After Koschmieder and Pallas (1974).

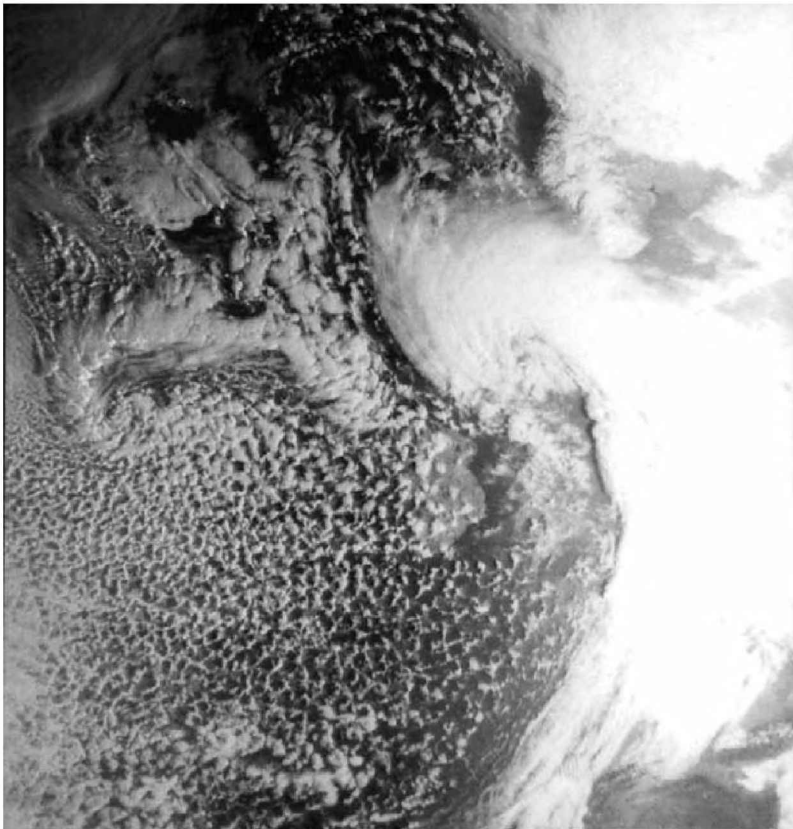
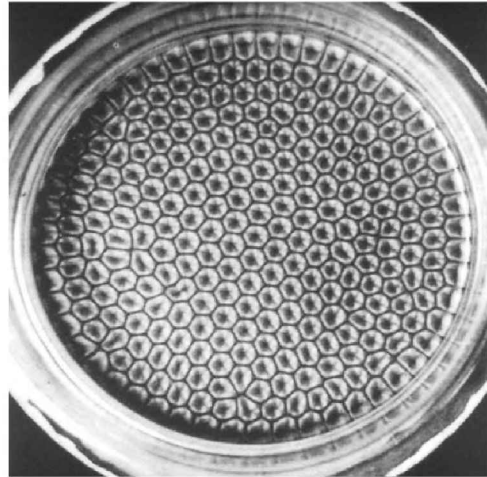


Figure 9.21 Visible image revealing cellular cloud structure behind a cold front. The cloud pattern is organized into a quasi-regular array of cells with a preferred horizontal scale. The cold front is part of a mature cyclone that advances across the British isles. Cloud cover ahead of the cold front marks the warm sector of the cyclone, which is just occluding north of Ireland. After Scorer (1986).

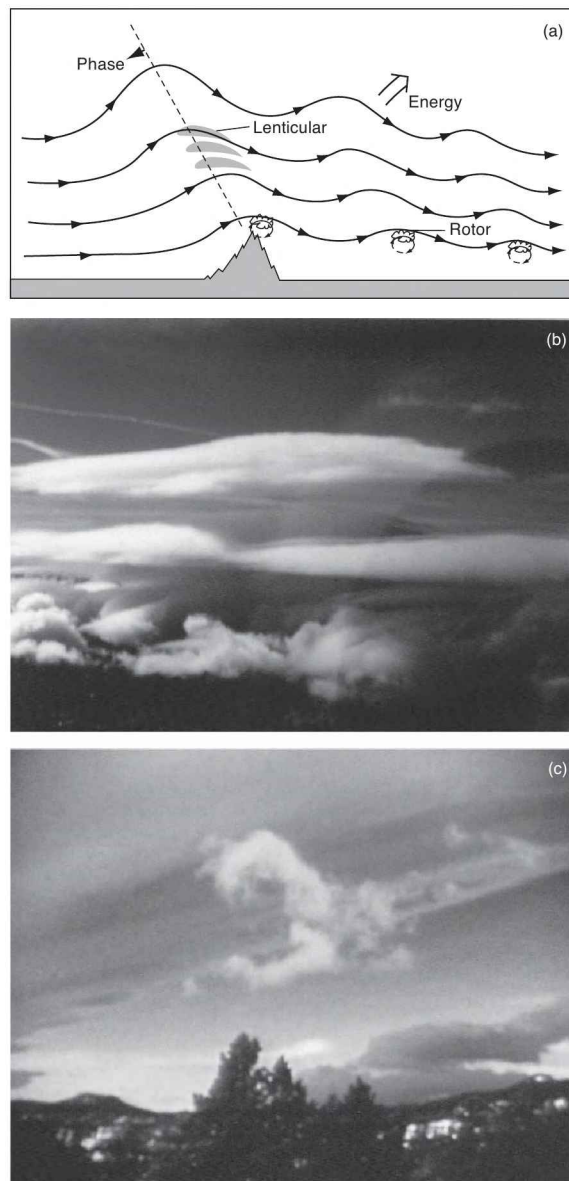


Figure 9.22 (a) Schematic of the mountain wave that develops leeward of an elongated range. *Lenticular clouds* are fixed with respect to orography, while air flows through them. The positions of *rotor clouds* are also fixed, but those fragmented features are highly unsteady, with turbulent ascent (descent) along their leading (trailing) edges. Westward tilt with height (e.g., of wave crests) corresponds to upward propagation of energy (Chap. 14), a feature that is often visible in the structure of lenticular clouds. (b) Longitudinal view through the mountain wave. Smooth lenticular clouds form in a stack, separated by clear regions, due to stratification of moisture. These laminar features contrast sharply with turbulent rotor clouds below. (c) Transverse view from the ground. Vorticity and severe turbulence are made visible by looping cloud matter that comprises rotor. Courtesy of V. Haynes and L. Feierabend (Soaring Society of Boulder). See color plate section: Plate 6.

made partially visible by *fractus* (fragmented) *cloud*. So-called *rotor cloud* is evident at levels beneath the lenticular in Figs. 9.22a,b. It is “dynamic.” Looping cloud matter in Fig. 9.22c makes visible strong updrafts and downdrafts. Severe turbulence accompanying that vertical motion has been responsible for numerous aircraft disasters over mountainous terrain.

Because orographic cloud remains fixed, it makes cloud behavior unreliable as a tracer of air motion. More faithfully tracking the movement of individual bodies of air is cumulus cloud. However, even it is complicated by nonconservative behavior (e.g., condensation and evaporation), which limits its usefulness for diagnosing air motion.

The foregoing cloud forms develop through upward motion. Ascent reduces the saturation mixing ratio of air through adiabatic cooling. Cloud can also develop isobarically through heat rejection, which cools air below its dew point. Conductive and radiative cooling lead to the formation of fog and marine stratus over cold ocean.

The stratosphere is isolated from convective motion and therefore from the source of water vapor at the Earth’s surface. Consequently, cloud rarely forms above the tropopause. Nacreous cloud is an exception. It develops through the mountain wave, which can propagate into the stratosphere (Chap 14). Nacreous cloud is a subset of more ubiquitous Polar Stratospheric Cloud (PSC). Because the stratosphere is very dry, PSC forms only under very cold conditions. Temperature over the Antarctic decreases below 190 K during Austral winter, more than 20K colder than temperature over the wintertime Arctic (Chap. 18). For this reason, PSC occurs over the Antarctic with much greater frequency and depth than over the Arctic.

Even over Antarctica, PSC is tenuous by comparison with tropospheric cloud. Some 3–5 km thick, PSC forms at heights of 15–25 km. It is often layered. Formation of PSC is strongly correlated with temperature. PSC is seldom sighted at temperatures warmer than 195 K. Conversely, its likelihood approaches 100% at temperatures below 185 K (WMO, 1988). Two distinct classes of PSC are observed. Type I PSC is composed of frozen droplets of nitric acid trihydrate (NAT), which has a low saturation vapor pressure. Submicron in scale, PSC I particles form at temperatures of 195–190 K. They have the appearance of haze. Type II PSC may be nucleated by Type I particles or by background aerosol. It contains much larger crystals of water ice. PSC II particles are several tens of microns and larger. Forming at temperatures of 185–190 K, they have features similar to cirrus. Their larger sizes give type II particles significant fall speeds. This feature makes them important in chemical considerations (Chap. 18).

9.3.2 Microphysical properties of cloud

Microphysical properties vary with cloud type (Table 9.3). Stratus has number densities $n \sim 300 \text{ cm}^{-3}$ and droplet radii of $a \sim 4 \mu\text{m}$. Liquid water content is more variable, but of order 0.5 g m^{-3} . Similar numbers apply for cumulus. However, cumulonimbus is distinguished by much smaller number density and much greater liquid water content. Implied are droplets that are significantly larger. Compared with water cloud, cirrus has much smaller number densities; they are observed over a wide range: $10^{-7} - 10 \text{ cm}^{-3}$ (Dowling and Radke, 1990; Knollenberg et al., 1993). Its ice water content ρ_i is likewise smaller (e.g., $\rho_i < 0.1 \text{ g m}^{-3}$). The characteristic dimension of cirrus ice crystals ranges from several tens of microns to as large as 1 mm. PSC has number densities similar to cirrus. It forms under analogous environmental conditions.

Table 9.3. Microphysical properties of cloud[%]

Cloud Type	n (cm ⁻³)	$\langle a \rangle$ (μm)	ρ_l (g m ⁻³)	$\langle \hat{\sigma}_e \rangle$ @0.5 μm (μm^2)	$\langle \beta_e \rangle$ @0.5 μm (km ⁻¹)
Stratus (St)	300.	3.	0.15	450.	135.
Stratocumulus (Sc)	250.	5.	0.3	120.	30.
Nimbostratus (Ns)	300.	4.	0.4	400.	120.
Cumulus (Cu)	300.	4.	0.5	200.	60.
Cumulonimbus (Cb)	75.	5.	2.5	500.	38.
Cirrus (Ci)	0.03	250.	0.025*	1.10 ⁴	0.5
Tropical Cirrus (Cs)	0.10	800.	0.20*	4.10 ⁴	0.4
PSC I	≤ 1	0.5	10 ^{-6*}	10	0.005
PSC II	≤ 0.10	50.	0.05*	100	0.015

[%] Representative values of number density n , mean droplet radius $\langle a \rangle$, liquid water content ρ_l , mean extinction cross section $\langle \hat{\sigma}_e \rangle$, and mean extinction coefficient $\langle \beta_e \rangle$ for different cloud types. Typical values are subject to a wide range of variability, especially for cirrus. *Sources:* Carrier et al. (1967), WMO (1988), Dowling and Radke (1990), Liou (1990), Knollenberg et al. (1993), Turco et al. (1993).

* Ice water content.

Figure 9.23 presents size spectra representative of several cloud types. Stratus and cumulus have the simplest spectra, which possess a single maximum. Spectra of nimbostratus and cumulonimbus exhibit secondary maxima, at radii larger than the mean: $\langle a \rangle \cong 5 \mu\text{m}$. The wings of these size spectra ($a > 20 \mu\text{m}$), are comparatively unpopulated. Nevertheless, they describe particles that grow most rapidly through collision and consequently favor precipitation (Sec. 9.2.2). Spectra of cirrus can also possess a secondary maximum, but at much larger size (e.g., at 100 – 1000 μm). Those larger particles have rapid fall speeds. They produce mares' tails, characteristic of hook-shaped *cirrus uncinus* (Fig. 9.15).

9.3.3 Cloud dissipation

Cloud dissolves through evaporation when unsaturated environmental air is entrained into and mixed with saturated air. By diluting high θ_e and momentum inside an updraft with lower values from the environment, entrainment siphons off positive buoyancy and kinetic energy. Affected likewise is remaining water vapor, along with the attendant latent heat that would otherwise be available to reinforce upward motion (Sec. 7.4.2). These features make cumulus entrainment the principal dissipation mechanism for cumulus. Cloud growth can continue only if it is offset by a resupply of sensible and latent heat (high θ_e) from below.

When particles become so large that they can no longer be supported by the updraft (e.g., $a > 100 \mu\text{m}$), cloud is dissipated by precipitation or, in the case of cirrus, sedimentation. Falling cloud particles exert downward drag on the air through which they descend. If they dissolve, they also subject that air to evaporative cooling. Both influences oppose the updraft. In doing so, they inhibit the supply of sensible and latent heat that is necessary to maintain the cloud against entrainment (Prob. 9.17).

Entrainment is less effective in stratiform cloud because of its horizontal extent and weaker vertical velocities. However, radiative heating at its base and cooling at

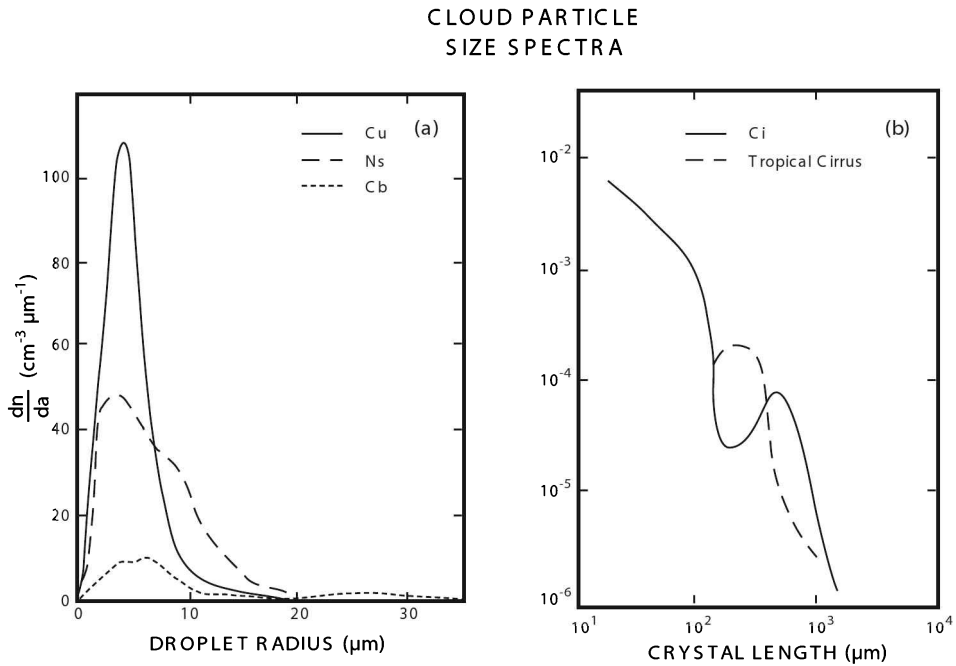


Figure 9.23 Cloud particle size spectra for (a) droplets inside water clouds and (b) crystals inside ice clouds. Adapted from Liou (1990).

its top destabilize stratus. This stimulates entrainment of drier air overhead, which acts to dissolve layered cloud. Radiative heating can also dissolve cloud by elevating its temperature above the dew point. Such heating leads to daily burn off of fog and marine stratus. This dissipation mechanism, together with formation mechanisms associated with surface heating, make diurnal variations an important feature of the cloud field (cf. Fig. 9.39).

9.3.4 Cumulus detrainment: Influence on the environment

At the wall of a cumulus updraft is entrainment of environmental air (Fig. 5.4). It is accompanied by detrainment of updraft air, which ventilates the environment. Thus, while it dilutes the θ_e of cumulus updrafts, this process simultaneously drives the environment toward the vertical structure inside updrafts (Sec. 7.7). The impact of cumulus detrainment is amplified in the upper troposphere. There, updrafts decelerate and diverge horizontally, forming extensive cirrostratus anvil.

Figure 9.24 presents a histogram of the fractional coverage by cloud, or the *cloud fraction* η . It is plotted as a function of brightness temperature, which translates into height. Representative of the tropics, η decreases upward exponentially, to less than 10% in the upper troposphere. Its sharp decrease with height reflects the histogram of updraft widths: Only the broadest updrafts escape damping by entrainment and, even then, only inside their cores. η levels off at temperatures colder than 235 K,

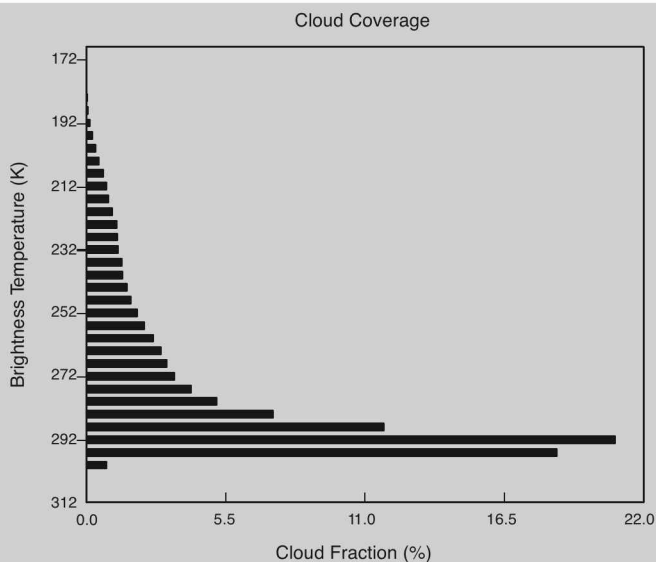


Figure 9.24 Histogram of fractional coverage by cloud, as a function of brightness temperature, which translates into altitude. After Salby et al. (2004a).

corresponding to ~ 250 hPa. It then decreases sharply at temperatures colder than 220 K, corresponding to levels above 200 hPa.⁷

Even though cloud fraction decreases upward monotonically, the coverage by individual anvils does not. Plotted in Fig. 9.25, also as a function of brightness temperature, is the mean areal extent of contiguous anvil during the same period as that presented in Fig. 7.14. Over South America (Fig. 9.25a), $\theta_{LNB} \cong 345$ K. This positions the LNB near 200 hPa, which is representative of much of the tropics (Sec 7.7). The extent of anvil increases upward, maximizing at brightness temperatures of 230–210 K. Corresponding to the layer 250–150 hPa, the maximum coincides with the LNB. It also coincides with the plateau in η , above which cloud fraction decreases sharply (Fig. 9.24). Over the equatorial Pacific (Fig. 9.25b), θ_{LNB} is 5–10 K warmer. This positions the LNB some 3 km higher than elsewhere in the tropics. There too, the extent of anvil increases upward. However, it now maximizes at a brightness temperature of 200 K, some 3 km higher.

In each region, the areal extent of anvil maximizes at levels neighboring the local LNB. The maximum marks extensive cirrostratus anvil. That cloud structure forms through detrainment of cumulus updrafts in the ITCZ. A small fraction of such air, inside the cores of broad updrafts, overshoots the LNB. It does so through kinetic energy that it acquired above the LFC (Fig. 7.6). After mixing with air overhead, it splashes back into an extensive anvil deck at the LNB, where updraft air diverges horizontally (Fig. 7.7).

⁷ Although rare, some cloud is observed at brightness temperatures as cold as 180°K. It reflects overshoots above the tropical tropopause, where environmental temperature is $\sim 200^\circ\text{K}$ (Fig. 1.7).

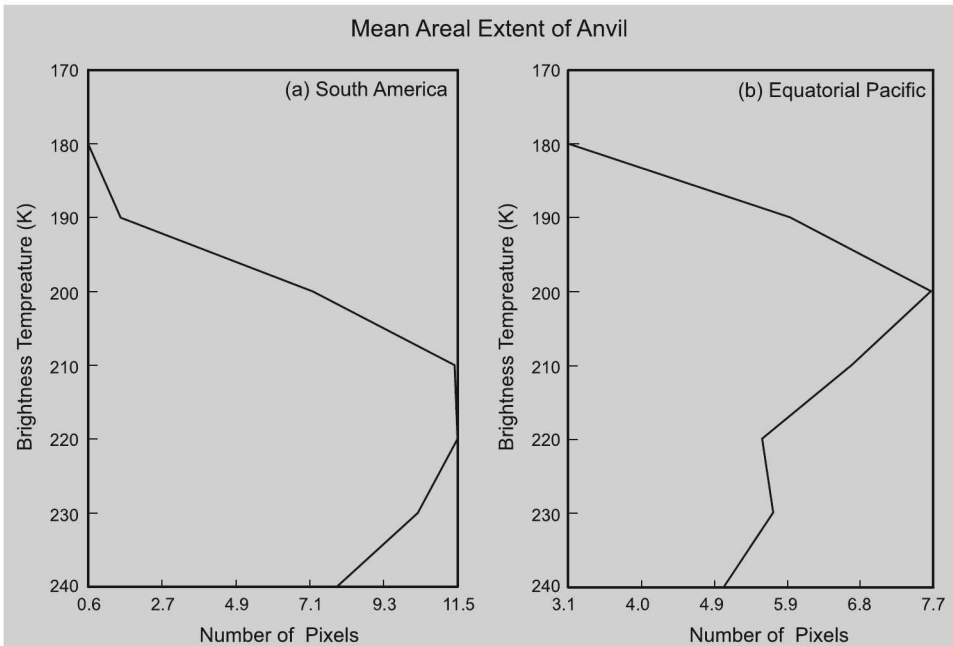


Figure 9.25 Mean areal extent of contiguous anvil (in number of 0.5° pixels), as a function of brightness temperature, for the same period as Fig. 7.14. After Salby et al. (2003).

In addition to cloud structure, such behavior is also manifest in the vertical structure of horizontal motion. Velocity divergence measures the collective outflow from deep cumulus towers (Sec. 10.5). Plotted in Fig. 9.26a for the equatorial Pacific, horizontal divergence maximizes at the same levels as extensive anvil. Air that diverges from cumulus updrafts in the ITCZ forms poleward motion. The latter comprises the upper branch of the Hadley circulation (Fig. 1.35). By supplying total water to an expansive area, that air also humidifies the layer of strong detrainment. Above the tropical tropopause, horizontal divergence reverses sign. Representing convergence, it implies weak downwelling of environmental air at lower levels, just above the layer of strong detrainment. The implied vertical motion is consistent with cumulus detrainment above the LNB, where it serves as a heat sink for the environment (Sec. 7.7).

Plotted in Fig. 9.26b for the same location and period is anomalous environmental mixing ratio (deviation from its NS mean). As η at these levels is only a couple of percent (Fig. 9.24), anomalous r represents the humidity of regions that are chiefly cloud-free. In the upper troposphere, positive anomalous r reflects a source of environmental water vapor. It is consistent with detrainment of total water from cumulus updrafts, a process that humidifies the environment. The implied source maximizes near 200 hPa, within the layer of large anvil extent and magnified detrainment (Figs 9.25b, 9.26a).

Overlying the layer where anomalous mixing ratio maximizes is a layer in which environmental RH exceeds 80% (shaded). Coincident with strong outflow, it appears within the layer of anomalously cold temperature (Fig. 7.15). The latter, which is

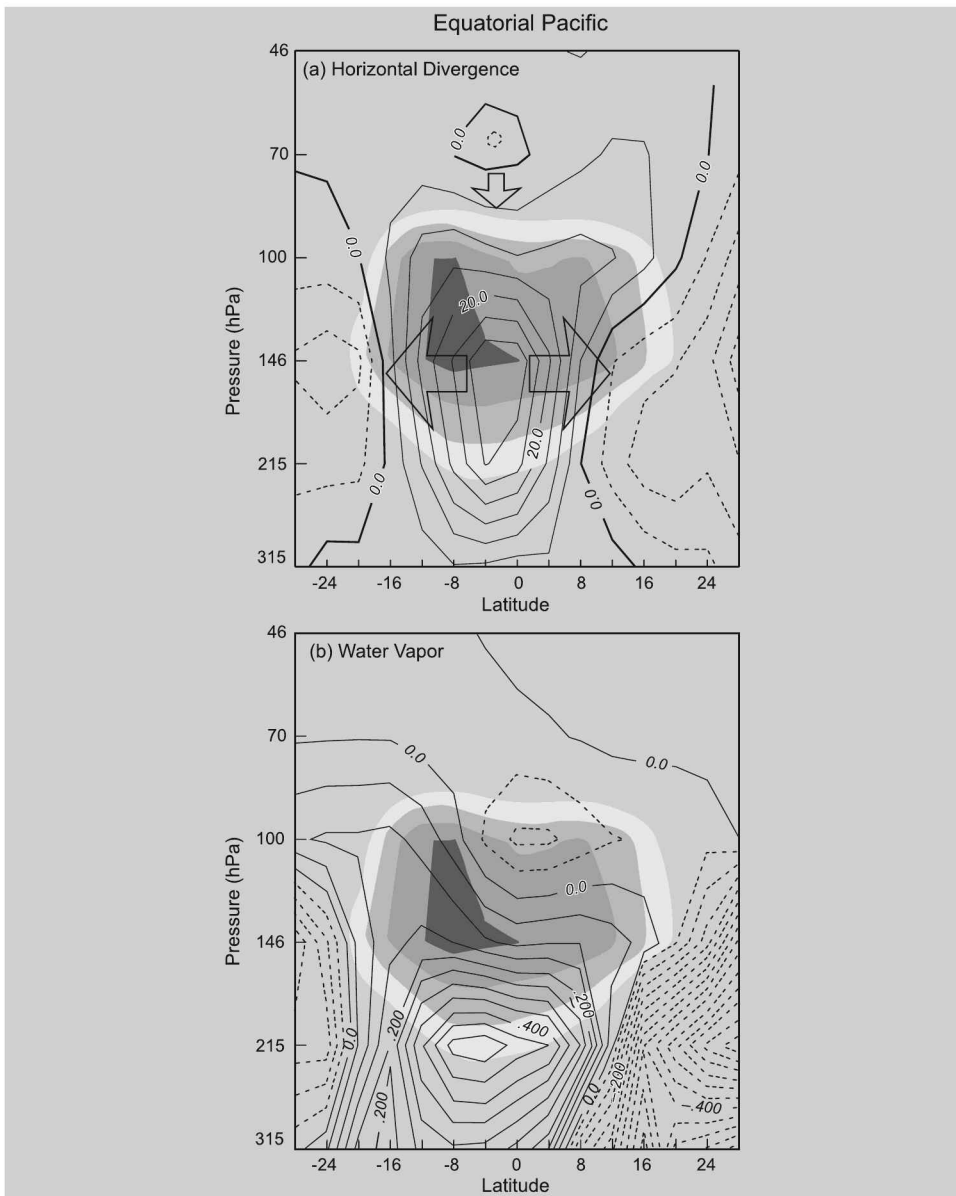


Figure 9.26 For the equatorial Pacific and the same period as Fig. 7.14, (a) horizontal velocity divergence, which measures the collective outflow from cumulus towers, and (b) anomalous humidity, departure from the NS mean of the logarithm of environmental mixing ratio. Shaded is RH exceeding 60%, 70%, and 80%. After Salby et al. (2003).

forced by cumulus overshoots, sharply reduces r_c . It thus drives environmental air toward saturation. These features of humidity structure are consistent with ice cloud being found preferentially at sites where RH is high (Wu et al, 2005). Inside cumulus overshoots, ice water content can be substantial. However, it involves large

particles, which are removed efficiently by precipitation to lower levels. Much the same applies to neighboring anvil.

Above the layer that is nearly saturated, anomalous mixing ratio reverses sign. Negative values at those levels reflect a sink of environmental water vapor. Their vertical position suggests a close relationship to the layer of nearly saturated conditions just below. Implied is efficient removal of total water as cumulus updrafts cross the layer of high RH. Because environmental air at those levels is nearly saturated, entrainment becomes ineffective at limiting condensation inside updrafts. Particle growth can then proceed unhindered, achieving large sizes that precipitate out (Danielsen, 1982). Depleted of total water, cumulus updrafts experiencing detrainment at higher levels would then dehydrate the environment.

9.4 RADIATIVE TRANSFER IN AEROSOL AND CLOUD

In the presence of liquid and solid particles, photons experience repeated reflection and diffraction. The path traveled by SW radiation is therefore greatly elongated. Supported by absorption inside particles, this sharply increases the optical depth posed to incident radiation.

9.4.1 Scattering by molecules and particles

Rayleigh scattering

The simplest treatment of scattering is due to Rayleigh (1871). It describes the interaction of sunlight with molecules. *Rayleigh scattering* applies to particles much smaller than the wavelength of radiation. It considers a molecule that is exposed to electromagnetic radiation as an oscillating dipole. The strength of this oscillation is measured by the *dipole moment* \mathbf{p} . It is related to the electric field \mathbf{E}_0 of incident radiation as

$$\mathbf{p} = \alpha \mathbf{E}_0, \quad (9.15)$$

where α is the *polarizability* of the scatterer. Established through interaction with the wave's electromagnetic field, the dipole in turn radiates a scattered wave with electric field \mathbf{E} . At large distance r from the dipole, radiation emitted at an angle γ from \mathbf{p} behaves as a plane wave whose amplitude satisfies

$$\mathbf{E} = \frac{1}{c^2} \frac{\sin\gamma}{r} \frac{\partial^2 \mathbf{p}}{\partial t^2}, \quad (9.16.1)$$

where c is the speed of light (e.g., Jackson, 1975). Incorporating (9.15) and plane wave radiation of the form $e^{ik(r-ct)}$ then obtains

$$\mathbf{E} = -\mathbf{E}_0 k^2 \alpha \frac{\sin\gamma}{r} e^{ik(r-ct)}. \quad (9.16.2)$$

Consider sunlight at a *scattering angle* Θ from the path of incident radiation (Fig. 9.27). Sunlight is unpolarized: Its electric field \mathbf{E}_0 is distributed isotropically over directions orthogonal to the path of propagation. It may therefore be represented in independent components \mathbf{E}_{0x} and \mathbf{E}_{0y} of equal magnitude. These incident electric fields induce dipole moments \mathbf{p}_x and \mathbf{p}_y , which in turn radiate scattered waves. Radiation scattered at the angle Θ makes an angle $\gamma_x = \frac{\pi}{2}$ from \mathbf{p}_x and an angle $\gamma_y = \frac{\pi}{2} - \Theta$

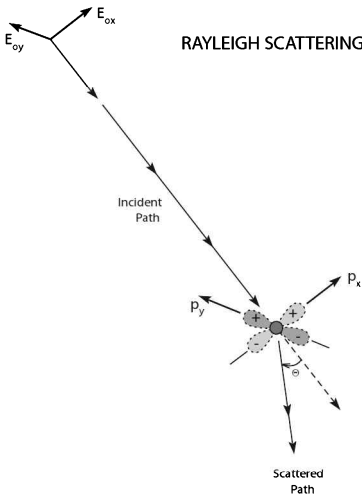


Figure 9.27 Rayleigh scattering of SW radiation by air molecules. Unpolarized sunlight is characterized by radiation with equal and independent electric fields \mathbf{E}_{0x} and \mathbf{E}_{0y} , which induce dipole moments \mathbf{p}_x and \mathbf{p}_y at the scatterer. Those in turn emit radiation at angles Θ from the path of incident radiation with electric fields \mathbf{E}_x and \mathbf{E}_y that are determined by \mathbf{p}_x and \mathbf{p}_y (see text). The orthogonal electric fields \mathbf{E}_{0x} and \mathbf{E}_{0y} have been chosen to position the scattering angle Θ in the plane formed by the path of incident radiation and \mathbf{E}_{0y} .

from \mathbf{p}_y . Then (9.16.2) gives the corresponding electric fields of scattered radiation

$$\begin{aligned} \mathbf{E}_x &= -\mathbf{E}_{0x} k^2 \alpha \frac{1}{r} e^{ik(r-ct)} \\ \mathbf{E}_y &= -\mathbf{E}_{0y} k^2 \alpha \frac{\cos\Theta}{r} e^{ik(r-ct)}. \end{aligned} \tag{9.17}$$

The intensity of radiation is proportional to the average of $|\mathbf{E}|^2$. Incident sunlight of intensity I_0 is then scattered into the angle Θ with intensity

$$I(\Theta, r) = I_0 k^4 \frac{\alpha^2}{r^2} \frac{(1 + \cos^2\Theta)}{2}. \tag{9.18}$$

In terms of the phase function (8.22) and wavelength $\lambda = \frac{2\pi}{k}$, (9.18) may be cast into the canonical form

$$I(\Theta, r) = I_0 \frac{\alpha^2}{r^2} \frac{32\pi^4}{3\lambda^4} P(\Theta), \tag{9.19.1}$$

with

$$P(\Theta) = \frac{3}{4} (1 + \cos^2\Theta). \tag{9.19.2}$$

The scattered intensity (Fig. 8.8a) has maxima in the forward ($\Theta = 0^\circ$) and backward ($\Theta = 180^\circ$) directions, with equal energy directed into each half-space.

The scattered flux follows as in (9.19), but with the incident flux F_0 in place of I_0 . Integrating the scattered flux over a sphere of radius r obtains the scattered power

$$\mathcal{P} = F_0 \alpha^2 \frac{128\pi^5}{3\lambda^4}, \tag{9.20}$$

which has dimensions of $\frac{\text{energy}}{\text{time}}$. The scattering cross section for an individual molecule is then

$$\begin{aligned} \hat{\sigma}_s &= \frac{\mathcal{P}}{F_0} \\ &= \alpha^2 \frac{128\pi^5}{3\lambda^4}. \end{aligned} \tag{9.21}$$

Finally, the scattered intensity at distance r may be expressed

$$I(\Theta, r) = I_0 \frac{\hat{\sigma}_s}{4\pi r^2} P(\Theta). \quad (9.22)$$

Electromagnetic field theory relates the polarizability α to the dimensionless refractive index

$$m = m_r - m_i. \quad (9.23)$$

The real and imaginary parts of m refer to the phase speed and absorption of electromagnetic radiation in a medium, relative to those in a vacuum. An ensemble of scatterers with number density n has polarizability

$$\alpha = \frac{3}{4\pi n} \left(\frac{m^2 - 1}{m^2 + 2} \right). \quad (9.24)$$

At wavelengths of visible radiation, absorption by air molecules is small enough for m_i to be ignored, whereas m_r is close to unity. Then (9.24) reduces to

$$\alpha \cong \frac{m_r - 1}{2\pi n}. \quad (9.25)$$

The scattering cross section for an individual molecule is thus given by

$$\hat{\sigma}_s = \frac{32\pi^3 (m_r - 1)^2}{3n^2 \lambda^4}. \quad (9.26)$$

Scattering then yields an optical depth for the atmosphere of

$$\tau = \hat{\sigma}_s(\lambda) \int_0^\infty n(z) dz. \quad (9.27)$$

Rayleigh scattering explains why the sky appears blue. The λ^{-4} dependence of $\hat{\sigma}_s$ implies that short wavelengths are scattered by air molecules much more effectively than long wavelengths. Blue light ($\lambda \sim 0.42 \mu\text{m}$) is scattered five times more than red light ($\lambda \sim 0.65 \mu\text{m}$). It therefore arrives at the Earth's surface as diffuse radiation, emanating from all directions. When the sun is near the horizon, sunlight passes obliquely through the atmosphere and therefore along an extended path. Shorter wavelengths are then scattered out of the incident beam, leaving only longer (red) wavelengths to illuminate the sky. On average, about 40% of the SW flux is scattered out of the incident beam in the near-UV. Conversely, less than 1% is removed in the near-IR. Overall, about 10% of SW radiation incident on the atmosphere is scattered by molecules. Half of it is returned to space, contributing to the Earth's albedo (Fig. 1.32). Most of that scattering takes place in the lowest 10 km, where $n(z)$ is large. For this reason, increasing elevation witnesses a darkening of the sky, along with an intensification and whitening of direct sunlight. Rayleigh scattering influences ozone photochemistry and stratospheric heating. It enhances the SW flux available to the Chappuis bands of O_3 (Sec. 8.3.1).

Owing to its relationship to dipole radiation, Rayleigh scattering has certain polarization properties. In the forward and backward directions ($\Theta = 0, 180^\circ$), scattered radiation is unpolarized - like incident sunlight. However, scattered sunlight becomes increasingly polarized at intermediate angles. It is fully polarized at directions orthogonal to the path of incident radiation ($\Theta = 90^\circ$). For this reason, the sky assumes a

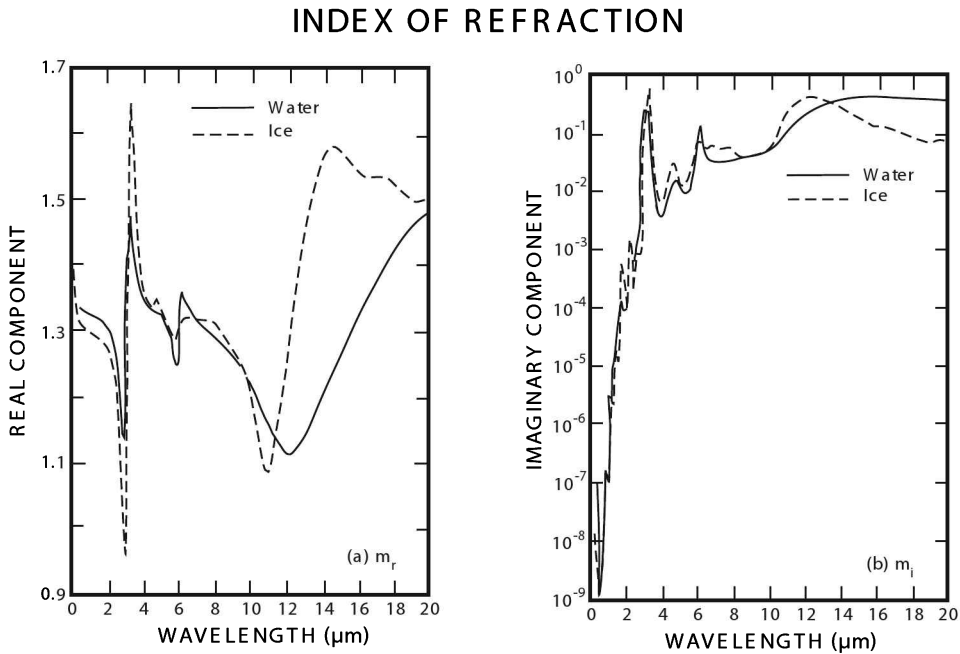


Figure 9.28 Index of refraction for water (solid) and ice (dashed) as functions of wavelength. Adapted from Liou (1990).

deeper blue when viewed through a polarizer at directions orthogonal to incident sunlight.

Mie scattering

Scattering from spherical particles of arbitrary dimension was first treated by Mie (1908). *Mie scattering* applies to the interaction of radiation with aerosol and cloud droplets. Figure 9.28 presents the refractive indices of water and ice, each as a function of wavelength. m_r is of order unity, but varies with λ . This feature makes diffraction and reflection by spheres of condensate wavelength-dependent. m_i is proportional to the absorption coefficient $\beta_{a\lambda}$:

$$m_i = \beta_{a\lambda} \frac{\lambda}{4\pi}. \tag{9.28}$$

It is small in the visible. However, m_i increases sharply at $\lambda > 1 \mu\text{m}$ for both condensed phases. Consequently all but tenuous cloud is optically thick in the IR.

Application of Mie scattering to atmospheric aerosol relies on particles being sufficiently separated for their interactions with the radiation field to be treated independently. Mie theory solves Maxwell’s equations for the electromagnetic field about a dielectric sphere of radius a in terms of an expansion in spherical harmonics and Bessel functions; see Liou (1980) for a detailed treatment. Properties of the scattered wave field emerge in terms of the dimensionless size parameter

$$x = 2\pi \frac{a}{\lambda}. \tag{9.29}$$

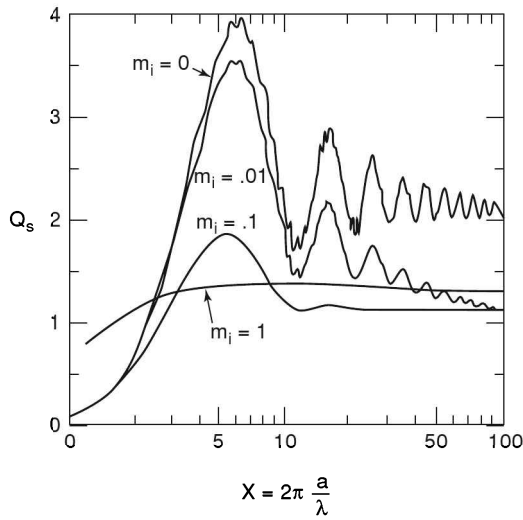


Figure 9.29 Scattering efficiency from a dielectric sphere having real component of refractive index $m_r = 1.33$ for several imaginary components m_i , as functions of the dimensionless size parameter $x = 2\pi \frac{a}{\lambda}$. After Hansen and Travis (1974). Reprinted by permission of Kluwer Academic Publishers.

The dimensionless *scattering* and *extinction efficiencies*

$$Q_s = \frac{\hat{\sigma}_s}{\pi a^2} \quad (9.30)$$

$$Q_e = \frac{\hat{k}}{\pi a^2}$$

represent the fractional area of the incident beam that is removed through interaction with the sphere by scattering and *in toto*. The scattering efficiency is shown in Fig. 9.29 for a sphere with $m_r = 1.33$ and for several values of m_i . For conservative scattering ($m_i = 0$), Q_s increases quartically for $x \ll 1$ corresponding to Rayleigh scattering. It attains a maximum near $x \cong 2\pi$ (i.e., at $\lambda \cong a$). The scattering efficiency then passes through a series of oscillations. They result from interference between light transmitted through and diffracted about the sphere.

In the limit of large radius, the oscillations in scattering efficiency damp out, yielding the limiting behavior

$$Q_s \sim 2 = \text{const} \quad x \rightarrow \infty \quad (9.31)$$

for conservative scattering ($m_i = 0$). Hence a sphere much larger than the wavelength of incident radiation scatters twice as much energy as it intercepts. Scattered radiation then includes contributions from energy that is diffracted about the sphere as well as energy that is redirected by reflection inside the sphere. This limiting behavior applies to the interaction of SW radiation with cloud droplets (cf. Fig. 9.23). Notice that, in this limit, Q_s depends only weakly on wavelength. Further, droplet size spectra span several of the weak oscillations that do remain. As a result, those oscillations experience cancellation. The scattering efficiency is therefore indiscriminate over wavelength. This feature of Mie scattering explains why cloud appears white.

For nonconservative scattering ($m_i > 0$), Q_s exhibits similar behavior. However, the oscillations due to interference are damped out with increasing absorption (because radiation emerges from the sphere progressively weaker). By $m_i = 1$, only the maximum near $\chi \cong 2\pi$ remains. Increased attenuation inside the sphere also reduces the limiting value of Q_s for $\chi \rightarrow \infty$, by absorbing energy that would otherwise contribute to the scattered wave field. Far from the sphere, radiation exhibits strong forward scattering (see Fig. 8.8b). However, for an individual sphere, it is complicated by rapid fluctuations with Θ that result from interference.

Under the conditions of independent scattering, properties of a population of spheres follow by collecting contributions from the individual particles. The scattering and extinction coefficients are then described by

$$\begin{aligned}\beta_s &= \int \hat{\sigma}_s dn(a) \\ \beta_e &= \int \hat{k} dn(a),\end{aligned}\tag{9.32}$$

where $dn(a) = \frac{dn}{da} da$ represents the droplet size spectrum. The single-scattering albedo for the population (8.20) is characterized by

$$\omega = \frac{\beta_s}{\beta_e}.\tag{9.33}$$

Similarly, the optical depth of a cloud of thickness Δz_c may be estimated as

$$\tau_c = \beta_e \Delta z_c,\tag{9.34}$$

where β_e is representative of the cloud as a whole.

Figure 9.30 shows, for several wavelengths, the phase function (solid) corresponding to a droplet size spectrum representative of cumulus (Fig. 9.23). Compared with Rayleigh scattering (dashed), Mie scattering possesses much stronger directionality. All λ shown exhibit strong forward scattering. This is especially true of short wavelengths, which are sharply peaked about $\Theta = 0$ and experience little absorption (Fig. 9.28b). A weaker maximum appears at backward scattering for wavelengths in the near-IR and visible.

The mean extinction cross section for the population is given by $\langle \sigma_e \rangle = \frac{\beta_e}{n}$ (Fig. 9.31a). It is nearly constant across the visible, for which cloud droplets are large (9.31). Weak absorption at those wavelengths leads to a single scattering albedo (Fig. 9.31b) of near unity. Hence $\sigma_e \cong \sigma_s$ and scattering is likewise wavelength independent. Individual components of sunlight are thus scattered with equal efficiency, so cloud appears white. The extinction cross section attains a maximum near $5 \mu\text{m}$, at a wavelength comparable to the mean radius of cloud droplets (Fig. 9.23). It decreases at longer wavelength. Because $\omega \sim 1$ for $\lambda < 10 \mu\text{m}$, most of the extinction results from scattering. At wavelengths of SW radiation, the latter dominates over absorption. Exceptional are narrow bands at $\lambda \cong 3 \mu\text{m}$ and $6 \mu\text{m}$, where absorption spectra of water and ice are peaked (Fig. 9.28b). At $\lambda > 10 \mu\text{m}$, ω decreases due to diminishing Q_s and increasing absorption. Similar optical properties follow from a droplet size spectrum representative of stratus.

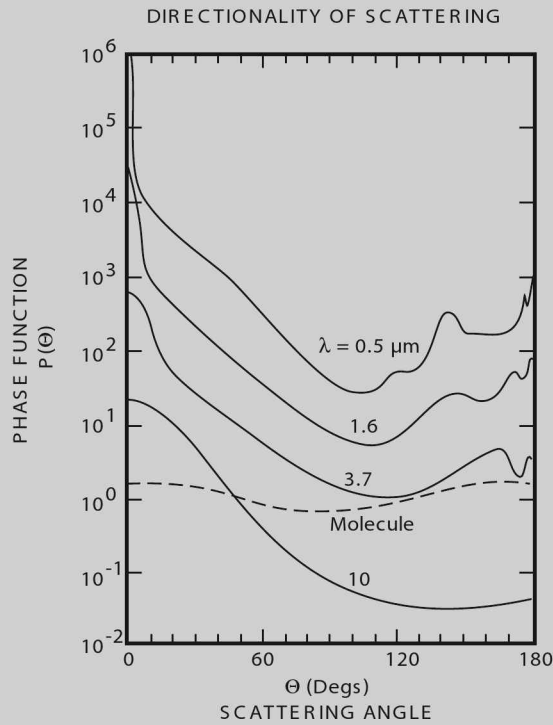


Figure 9.30 Phase function for scattering from the cumulus droplet population in Fig. 9.23, for several wavelengths (solid). Phase function for Rayleigh scattering (dashed) superposed. *Source:* Liou (1990).

CLOUD OPTICAL PROPERTIES

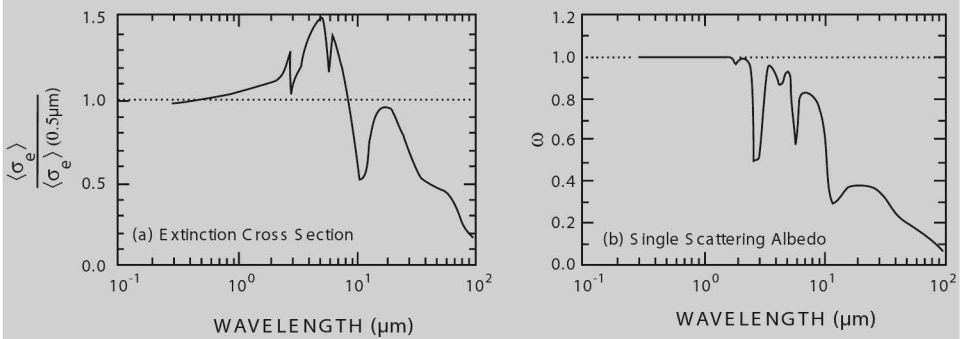


Figure 9.31 (a) Extinction cross section normalized by $(\sigma_e)(0.5 \mu\text{m}) = 166 \mu\text{m}^2$ and (b) single scattering albedo for the cumulus droplet population in Fig. 9.23. Adapted from Liou (1990).

Extinction by cloud droplets greatly increases the SW opacity of the atmosphere. For cumulus 1 km thick and a droplet number density of 300 cm^{-3} , (9.34) gives an optical depth at $\lambda = 0.5 \text{ }\mu\text{m}$ of $\tau_c = 50$. The cloud is optically thick to SW radiation. Organized cumulus of the same dimensions can have $\tau_c > 100$. Even shallow stratus has optical depths in the visible well in excess of 10. In the IR, τ_c is smaller. However, it is still large enough to render all but tenuous cloud optically thick to LW radiation. At $10 \text{ }\mu\text{m}$, the foregoing conditions for cumulus lead to an optical depth of about 20.

Scattering by nonspherical ice particles is complicated by their irregular structure and anisotropy. Even though their rudimentary geometry is hexagonal, ice crystals assume a wide range of shapes (Fig. 9.14). They can produce a myriad of optical effects associated with orientation, directionality, and polarization. The familiar halo of thin cirrus results from prismatic crystals with a preferred orientation.

9.4.2 Radiative transfer in a cloudy atmosphere

How cloud and aerosol influence atmospheric thermal structure is now evaluated through the radiative transfer equation. Consider a homogeneous scattering layer that is illuminated by a solar flux F_s , which is inclined at the zenith angle $\theta_s = \cos^{-1}\mu_s$ (Fig. 9.32). The source function for diffuse SW radiation (8.24) then becomes

$$J(\phi, \mu, \tau) = \frac{\omega}{4\pi} \int_0^{2\pi} \int_{-1}^1 I(\phi', \mu', \tau) P(\phi, \mu; \phi', \mu') d\phi' d\mu' + \frac{\omega}{4\pi} F_s P(\phi, \mu; \phi', -\mu_s) e^{-\frac{\tau}{\mu_s}}, \quad (9.35)$$

where the wavelength dependence is implicit and emission is ignored. The first term on the right-hand side represents contributions to diffuse radiation from multiple scattering of the diffuse intensity I . The second term represents the contribution from single scattering of direct solar radiation. Transmission of diffuse SW radiation through a cloudy atmosphere is then governed by the radiative transfer equation.

Scattering transforms (8.25) into an integro-differential equation. It is treated by expanding the phase function $P(\cos \Theta)$, where Θ is the 3-dimensional scattering angle, in a series of spherical harmonics and the diffuse intensity I in a like manner; see Liou (1980) for a formal treatment. Expressing $P(\cos \Theta)$ in terms of the spherical coordinates (ϕ, θ) and integrating over ϕ then yields the transfer equation governing the azimuthal-mean component $I(\mu, \tau)$ of the diffuse radiation field

$$\mu \frac{dI}{d\tau} = I - \frac{\omega}{2} \int_{-1}^1 I(\mu', \tau) P(\mu; \mu') d\mu' - \frac{\omega}{4\pi} F_s P(\mu; -\mu_s) e^{-\frac{\tau}{\mu_s}}, \quad (9.36.1)$$

where

$$P(\mu; \mu') = \sum_j c_j P_j(\mu) P_j(\mu'), \quad (9.36.2)$$

P_j is the Legendre polynomial of degree j , and the expansion coefficients c_j follow from orthogonality properties of the $P_j(\mu)$. For single scattering, the convolution integral is omitted. The diffuse intensity is then directly proportional to the phase function P . Approximate solutions for multiple scattering can be obtained by numerical techniques. A simple but enlightening solution follows from the two-stream approximation.

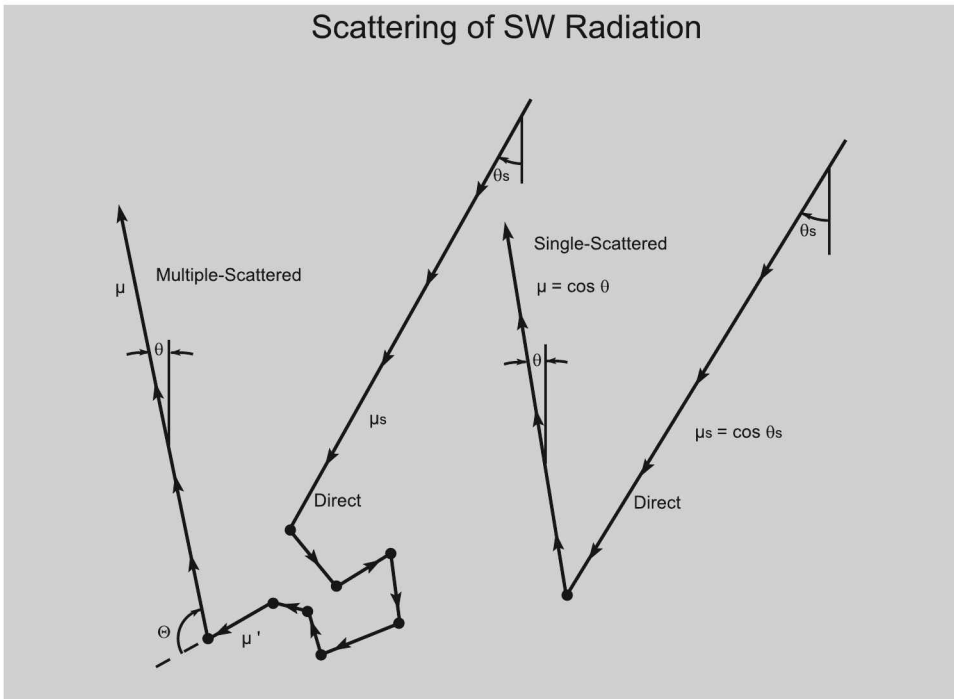


Figure 9.32 Transmission of diffuse SW radiation through a scattering layer.

Owing to the involvement of Legendre polynomials, the integral in (9.36) can be evaluated efficiently with Gaussian quadrature at a finite number of zenith angles μ_j (e.g., Rektorys, 1969). Each describes a different stream of radiation. The two-stream approximation then follows by taking just two angles: $j = \pm 1$, which correspond to the diffusivity factor $\bar{\mu}^{-1} = \sqrt{3}$, i.e., to the upwelling and downwelling streams $I^{\uparrow\downarrow} = I(\pm\bar{\mu}, \tau)$. Incorporating the orthogonality properties of Legendre polynomials (Abramowitz and Stegun, 1972) and the transformation (8.54) leads to a system of two ordinary differential equations for the downwelling and upwelling fluxes of diffuse SW radiation

$$-\frac{dF^\downarrow}{d\tau^*} = F^\downarrow - \omega [fF^\downarrow + (1 - f)F^\uparrow] - S^+ e^{-\frac{\rho}{\mu_s} \tau^*} \tag{9.37.1}$$

$$\frac{dF^\uparrow}{d\tau^*} = F^\uparrow - \omega [fF^\uparrow + (1 - f)F^\downarrow] - S^- e^{-\frac{\rho}{\mu_s} \tau^*}, \tag{9.37.2}$$

where τ^* is given by (8.54),

$$f = \frac{1 + g}{2} \tag{9.37.3}$$

and $(1 - f)$ represent the fractional energy that is forward and backward scattered, respectively,

$$g = \frac{1}{2} \int_{-1}^1 P(\cos\Theta) \cos\Theta d(\cos\Theta) \tag{9.37.4}$$

is an *asymmetry factor*, and

$$S^\pm = \omega \frac{F_s}{4} (1 \pm 3g\bar{\mu}\mu_s). \quad (9.37.5)$$

The asymmetry factor g equals the first moment of the phase function. It measures the difference in scattering between the forward and backward half spaces. For isotropic scattering, $g = 0$ and $f = \frac{1}{2}$. The same holds for Rayleigh scattering, which is symmetric. However, the strong forward lobe of Mie scattering (Figs. 8.8b, 9.30) leads to $f > \frac{1}{2}$, with $g \cong 0.85$ representative of water cloud.

Subtracting and adding (9.36), introducing the net and total fluxes

$$F = F^\uparrow - F^\downarrow \quad (9.38.1)$$

$$\bar{F} = F^\uparrow + F^\downarrow, \quad (9.38.2)$$

and differentiating with respect to τ^* results in the equivalent 2nd order system

$$\frac{d^2 F}{d\tau^{*2}} - \gamma^2 F = Z e^{-\frac{\bar{\mu}}{\mu_s} \tau^*} \quad (9.39.1)$$

$$\frac{d^2 \bar{F}}{d\tau^{*2}} - \gamma^2 \bar{F} = \bar{Z} e^{-\frac{\bar{\mu}}{\mu_s} \tau^*}, \quad (9.39.2)$$

where

$$\gamma^2 = (1 - \omega)(1 + \omega - 2\omega f) \quad (9.39.3)$$

$$Z = \frac{\bar{\mu}}{\mu_s} \bar{S} + (1 - \omega) S \quad (9.39.4)$$

$$\bar{Z} = -\frac{\bar{\mu}}{\mu_s} S - (1 + \omega - 2\omega f) \bar{S}$$

and

$$S = S^+ - S^- \quad (9.39.5)$$

$$\bar{S} = S^+ + S^-.$$

The system (9.39) governs the transmission of diffuse SW radiation. It has general solution

$$F^\uparrow = C\alpha_+ e^{\gamma\tau^*} + D\alpha_- e^{-\gamma\tau^*} + E_+ e^{-\frac{\bar{\mu}}{\mu_s} \tau^*} \quad (9.40.1)$$

$$F^\downarrow = C\alpha_- e^{\gamma\tau^*} + D\alpha_+ e^{-\gamma\tau^*} + E_- e^{-\frac{\bar{\mu}}{\mu_s} \tau^*},$$

where

$$\alpha_\pm = \frac{1 \pm \beta}{2}$$

$$\beta^2 = \frac{1 - \omega}{1 + \omega - 2\omega f}$$

$$E_\pm = \frac{\bar{\delta} \pm \delta}{2} \quad (9.40.2)$$

$$\bar{\delta} = \frac{\mu_s^2}{\bar{\mu}^2 - \mu_s^2 \gamma^2} \bar{Z} \quad \delta = \frac{\mu_s^2}{\bar{\mu}^2 - \mu_s^2 \gamma^2} Z.$$

The integration constants C and D are determined by boundary conditions. τ^* is measured from the top of the cloud layer, which has an effective optical depth τ_c^* . Likewise, the underlying surface is black. Then the downwelling diffuse flux vanishes at the top of the scattering layer, whereas the upwelling diffuse flux vanishes at the bottom:

$$\begin{aligned} F^\downarrow(0) &= 0 \\ F^\uparrow(\tau_c^*) &= 0. \end{aligned} \quad (9.41)$$

The integration constants then become

$$\begin{aligned} C &= \frac{E_- \alpha_- e^{-\gamma \tau_c^*} - E_+ \alpha_+ e^{-\frac{\beta}{\mu_s} \tau_c^*}}{\alpha_+^2 e^{\gamma \tau_c^*} - \alpha_-^2 e^{-\gamma \tau_c^*}} \\ D &= \frac{E_+ \alpha_- e^{-\frac{\beta}{\mu_s} \tau_c^*} - E_- \alpha_+ e^{\gamma \tau_c^*}}{\alpha_+^2 e^{\gamma \tau_c^*} - \alpha_-^2 e^{-\gamma \tau_c^*}}. \end{aligned} \quad (9.42)$$

The solution can be used to evaluate the cloud albedo and transmissivity

$$\mathcal{A}_c = \frac{F^\uparrow(0)}{\mu_s F_s} \quad (9.43.1)$$

$$\mathcal{T}_c = \frac{F^\downarrow(\tau_c^*) + \mu_s F_s e^{-\frac{\beta}{\mu_s} \tau_c^*}}{\mu_s F_s}. \quad (9.43.2)$$

In (9.43.2), the direct-solar contribution becomes negligible for cloud that is optically thick. The cloud absorptivity follows as $1 - \mathcal{A}_c - \mathcal{T}_c$.

Figure 9.33 plots the cloud albedo and transmissivity, each as a function of optical depth and solar zenith angle; values are for microphysical properties representative of cumulus (Fig. 9.23) at a wavelength of $0.5 \mu\text{m}$. Cloud albedo increases sharply at small optical depth (Fig. 9.33a). It exceeds 50% for $\tau_c > 20$, typical of water cloud. Hence, even shallow stratus are highly reflective. Transmissivity decreases with increasing τ_c - nearly complementary to cloud albedo. Thus, most incident SW energy is either reflected or transmitted. The absorptivity, $1 - \mathcal{A}_c - \mathcal{T}_c$, is of order 10% across much of the range of optical depth shown. Absorption by droplets and surrounding vapor is comparable. Because the cloud layer is optically thick, most of the absorption occurs near its top. There, the heating rate $\frac{q}{c}$ reaches 10 Kday^{-1} and greater. Cloud albedo also increases with increasing solar zenith angle (Fig. 9.33b), which elongates the slant optical path. For $\tau_c = 20$, \mathcal{A}_c increases from 53% for overhead sun to 75% when the sun is on the horizon. The behavior of \mathcal{T}_c is nearly complementary. It decreases from 32% for overhead sun to only 10% when the sun is on the horizon. Shallower cloud depends even more strongly on solar zenith angle because the slant optical path then varies between optically thin and optically thick conditions.

The factor that influences cloud optical properties most strongly is liquid water content. It determines τ_c . The optical depth of a cloud may be expressed in terms of the column abundance of liquid water,

$$\begin{aligned} \Sigma_l &= \int_0^\infty \rho_l dz \\ &= \Delta z_c \cdot \frac{4\pi \rho_w}{3} \int a^3 dn(a), \end{aligned} \quad (9.44)$$

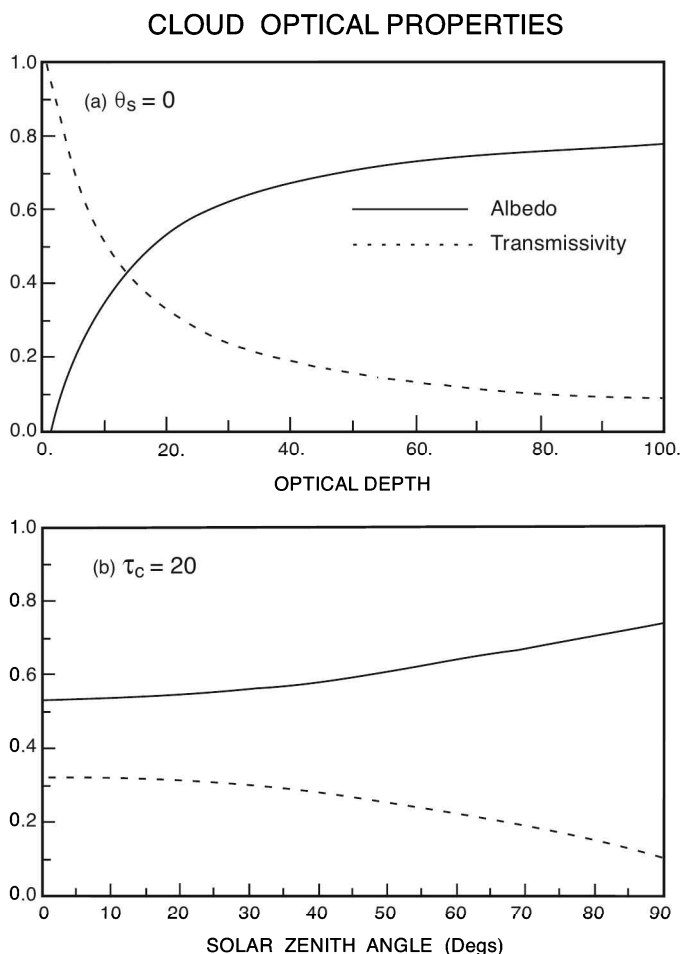


Figure 9.33 Albedo and transmissivity at $0.5 \mu\text{m}$ for a cloud layer with microphysical properties of cumulus in Fig. 9.23 and nearly conservative scattering (a) for overhead sun, as functions of cloud optical depth and (b) for a cloud optical depth of 20, as functions of solar zenith angle.

which is called the *liquid water path*. In the limit of large particles, which applies to SW radiation, the extinction cross section asymptotically approaches $\hat{\sigma}_e = Q_e \pi a^2 \cong 2\pi a^2$ (Fig. 9.29). The optical depth may then be expressed

$$\tau_c = \frac{3\Sigma_l}{2\rho_w a_e}, \quad (9.45.1)$$

where

$$a_e = \frac{\int a(\pi a^2) dn(a)}{\int (\pi a^2) dn(a)} \quad (9.45.2)$$

is a scattering-equivalent mean droplet radius. Most water cloud has liquid water content exceeding 0.1 g m^{-3} (Table 9.3). Therefore, cloud that is more than a kilometer thick has liquid water paths Σ_l of 100 g m^{-2} and greater. Such cloud then has optical depth in the visible of several tens and greater. As displayed in Fig. 9.34, this leads to

CLOUD OPTICAL PROPERTIES

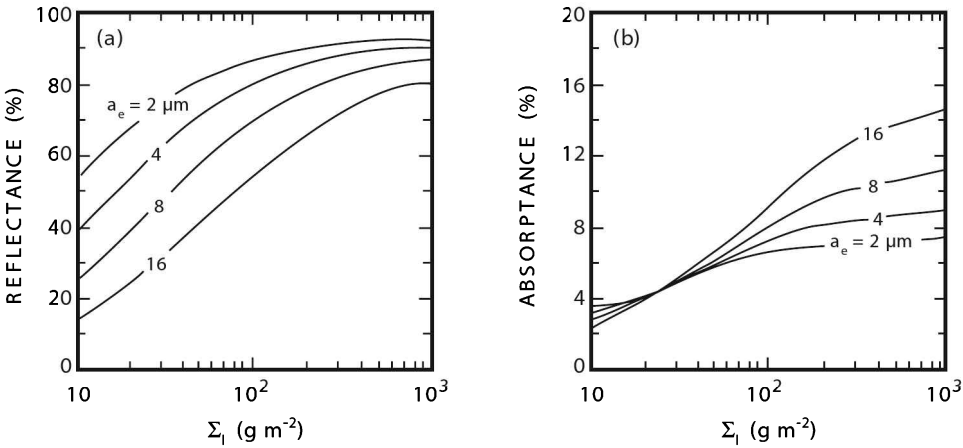


Figure 9.34 SW reflectance (a) and absorbance (b) as functions of liquid water path, for a solar zenith angle of 60° and for different scattering-equivalent droplet radii a_e . Adapted from Slingo (1989).

albedo of 50% and greater. Deep cloud with $\Sigma_l > 1000 \text{ g m}^{-2}$ has albedo that exceeds 80% (cf. bright features in Fig. 1.29).

For given Σ_l , cloud with small droplets has a higher albedo and smaller absorptivity than cloud with large droplets (Fig. 9.34). The increased reflectivity of cloud containing small droplets follows from the increased number densities in such cloud and the reduced absorption of small droplets. Its smaller droplets and greater number densities make continental cloud more reflective than maritime cloud of the same type and liquid water path (e.g., Twomey, 1977). For the same reason, precipitation ($a > 100 \mu\text{m}$) sharply increases the absorptivity of cloud. It is for this reason that precipitating cloud appears dark. Ice cloud, although more complex, is likewise highly reflective across the visible. Its dependence on microphysical properties resembles that of water cloud. Thin cirrus has ice water paths of $\Sigma_i < 5 \text{ g m}^{-2}$. It has albedo of order 10% and absorptivity of only a couple of percent (Paltridge and Platt, 1981).

According to Fig. 9.31b, scattering by cloud particles falls off at $\lambda > 4 \mu\text{m}$, where water absorbs strongly (Fig. 9.28b). Figure 9.35 illustrates that water cloud 3 km thick, which is highly reflective at $\lambda < 1 \mu\text{m}$, becomes fully absorbing at $\lambda > 4 \mu\text{m}$.

The foregoing development governs the transmission through cloud of SW radiation. Transmission of LW radiation can be treated in similar fashion if the scattering of direct solar in (9.37) is replaced by the emission of terrestrial radiation: $(1 - \omega)B^*[T]$. Inside a cloud layer of temperature T_c , upwelling and downwelling LW fluxes are then described by

$$\frac{dF^\uparrow}{d\tau^*} = F^\uparrow - \omega[fF^\uparrow + (1 - f)F^\downarrow] - (1 - \omega)B^*[T_c] \quad (9.46.1)$$

$$-\frac{dF^\downarrow}{d\tau^*} = F^\downarrow - \omega[fF^\downarrow + (1 - f)F^\uparrow] - (1 - \omega)B^*[T_c], \quad (9.46.2)$$

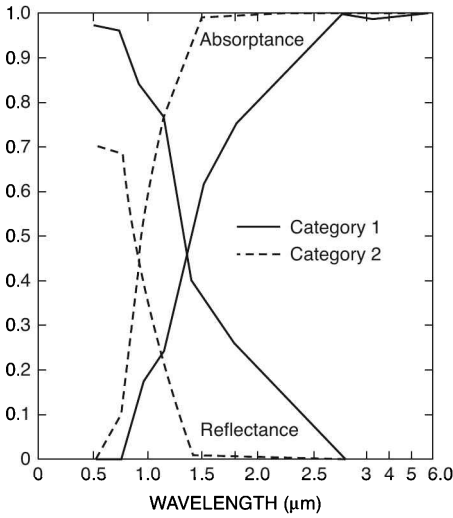


Figure 9.35 Reflectance and absorptance of 3-km thick stratus, as functions of wavelength for overhead sun and for a range of microphysical properties bounded by categories 1 and 2. Category 1 is based on a droplet distribution of characteristic radius $6 \mu\text{m}$, $n = 10^2 \text{ cm}^{-3}$, and $\rho_l = 0.30 \text{ g m}^{-3}$. Category 2 includes rain with a droplet distribution of characteristic radius $600 \mu\text{m}$, $n = 0.001 \text{ cm}^{-3}$, and $\rho_l = 2.1 \text{ g m}^{-3}$. Source: Cox (1981).

where the cloud is sufficiently thin for T_c to be treated as constant. A particular solution of (9.46) is the blackbody spectrum B^* . Proceeding as in the treatment of SW radiation leads to the general solution

$$\begin{aligned} F^\uparrow &= C\alpha_+ e^{\gamma\tau^*} + D\alpha_- e^{-\gamma\tau^*} + (1 - \omega)B^*[T_c] \\ F^\downarrow &= C\alpha_- e^{\gamma\tau^*} + D\alpha_+ e^{-\gamma\tau^*} + (1 - \omega)B^*[T_c], \end{aligned} \tag{9.47}$$

which applies inside the scattering layer.

To illustrate the influence of cloud on the LW energy budget, we consider a statistically homogeneous scattering layer within the framework of radiative-convective equilibrium (Sec. 8.5.2). Microphysical properties incorporated in the treatment of SW radiation (Fig. 9.23) are used to define a cloud layer of thickness $\Delta z_c = 1 \text{ km}$ and fractional coverage η that is situated at a height z_c in the troposphere - all under the conditions of radiative-convective equilibrium (Fig. 8.21). In the spirit of a gray atmosphere, radiative characteristics at $10 \mu\text{m}$ are used to define the broadband transmission properties for LW radiation. The fractional cloud cover is adjusted to give an albedo of 30%. F^\uparrow and F^\downarrow are described by (8.70) outside the cloud layer and by (9.47) inside the cloud layer. Matching fluxes across the scattering layer then leads to behavior analogous to that under simple radiative-convective equilibrium (Fig. 8.22), except for the addition of the cloud layer.

Figure 9.36a shows upwelling and downwelling fluxes in the presence of a cloud layer of optical depth $\tau_c = 20$ that is situated at $z_c = 8 \text{ km}$. The equilibrium temperature structure (not shown) is identical to that in Fig. 8.21, except that temperatures beneath the cloud layer are uniformly warmer. The surface temperature T_s has increased from its cloud-free value of 285 K to 312 K. Below the cloud layer, upwelling and downwelling fluxes resemble those under cloud-free conditions (Fig. 8.22a), except likewise displaced to higher values. Inside the cloud layer, the profiles of F^\uparrow and F^\downarrow undergo

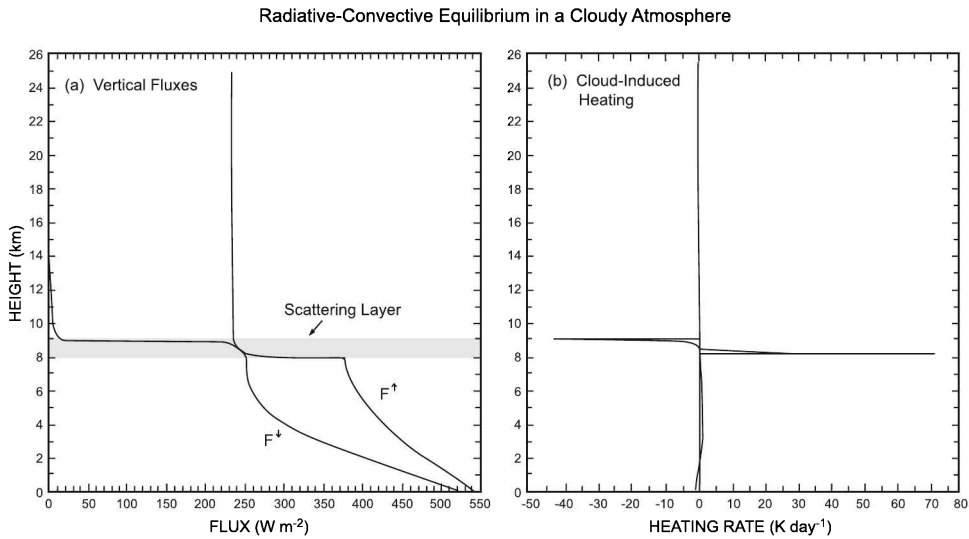


Figure 9.36 Energetics inside a cloudy atmosphere that is in radiative-convective equilibrium and contains a 1 km thick scattering layer with microphysical properties of cumulus in Fig. 9.23. The fractional cloud cover is adjusted to give an albedo of 0.30. (a) Upwelling and downwelling LW fluxes. (b) Cloud-induced heating: $\frac{\dot{q}}{c_p} - \left(\frac{\dot{q}}{c_p}\right)^{CS}$, where the superscript refers to radiative-convective equilibrium under clear-sky conditions (Fig. 8.22).

a sharp adjustment that drives them into coincidence. This behavior follows from the cloud being optically thick. The LW flux is then isotropic, equal to σT^4 . Above the cloud layer, F^\uparrow and F^\downarrow recover the cloud-free forms in Fig. 8.22a to satisfy boundary conditions at TOA (8.63).

Sharp changes of F^\uparrow and F^\downarrow in Fig. 9.36a introduce strong LW warming near the base of the cloud layer and strong LW cooling near its top. Figure 9.36b plots the cloud-induced heating: $\frac{\dot{q}}{c_p} - \left(\frac{\dot{q}}{c_p}\right)^{CS}$, where $\left(\frac{\dot{q}}{c_p}\right)^{CS}$ is the clear-sky heating profile under radiative-convective equilibrium (Fig. 8.22b). LW absorption (heating) is concentrated within an optical depth of the cloud base. LW emission (cooling) is concentrated within an optical depth of the cloud top. These regions of net absorption and emission are shallow, analogous to Chapman layers (Sec. 8.5.3). They introduce substantial heating and cooling locally (8.72). LW heating approaches 70 K day^{-1} near the cloud's base. LW cooling approaches 50 K day^{-1} near its top. By re-emitting energy downward, the cloud also introduces heating beneath its base. However, it is much weaker than the heating and cooling at the cloud's boundaries.

LW cooling in upper portions of a cloud is offset by SW heating, but only partially. Anvil cirrus experiences net radiative cooling at its top of several tens of K day^{-1} and heating at its base as large as 100 K day^{-1} (Platt et al., 1984a). Radiative cooling at its top and heating at its base destabilize a cloud layer. That, in turn, introduces convection, which entrains drier environmental air that tends to dissolve cloud (Fig. 9.19). Radiative heating also tends to dissolve cloud by elevating its saturation mixing ratio. Such behavior is typified by stratus burn off.

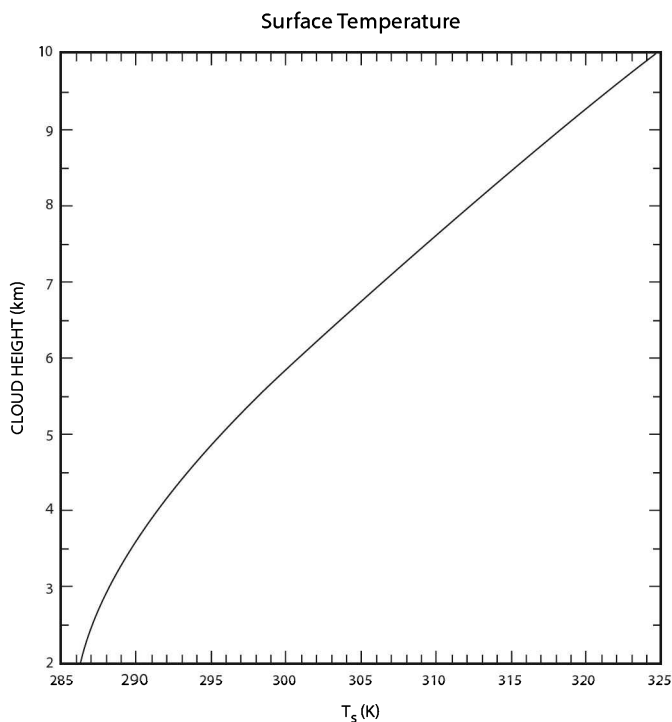


Figure 9.37 Surface temperature for the cloudy atmosphere in Fig. 9.36, as a function of the cloud layer's height.

The cloud layer in Fig. 9.36 is optically thick. Therefore, it influences the TOA LW energy budget chiefly through cloud-top emission. The latter decreases with the cloud's temperature and elevation. Figure 9.37 shows the dependence on cloud height of surface temperature under radiative-convective equilibrium. Situating the cloud below 2 km recovers nearly the same surface temperature as under cloud-free conditions. The scattering layer is then embedded inside the optically-thick layer of water vapor. It therefore has little influence on the effective emission temperature of the atmosphere. However, $z_c > 2$ km leads to warmer surface temperature. T_s increases steadily with the cloud's height. By $z_c = 8$ km, T_s has increased 27°K to maintain $F^\uparrow(\tau^* = 0) = (1 - \mathcal{A})\bar{F}_s$ and hence maintain the equivalent blackbody temperature of the Earth (Prob. 9.26).

Water cloud becomes optically thick to LW radiation for liquid water paths Σ_l greater than about 20 g m^{-2} . It then behaves as a blackbody. The LW characteristics of ice cloud are more complex, at least for cloud that is optically thin. The emissivity (LW absorptivity) of thin cirrus with ice water paths less than 10 g m^{-2} (typical of cirrus 1 km thick) is 0.5 or smaller. It can vary diurnally with convection (Platt et al., 1984b). However, even cirrus emits approximately as a blackbody for Δz_c sufficiently large. Figure 9.38 shows spectra of OLR in the tropics over a clear-sky region and several neighboring cirrus-covered regions of different optical depth. Cirrus reduces the radiation emitted to space throughout the LW, except near $15 \mu\text{m}$. There, radiation

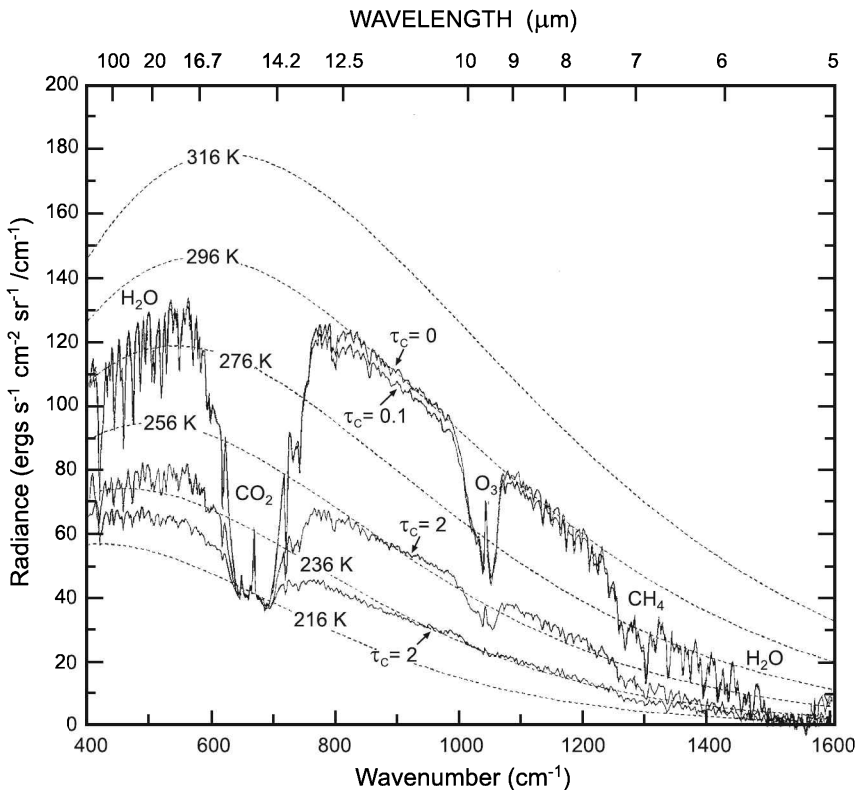


Figure 9.38 Spectrum of outgoing LW radiation observed by Nimbus-4 IRIS for a clear sky region ($\tau_c = 0$) at (134E, 12N) and neighboring cirrus-covered regions ($\tau_c > 0$). Courtesy of B. Carlson (NASA/Goddard Institute for Space Studies).

emitted by the surface has already been absorbed by underlying CO_2 . By an optical depth of $\tau_c = 5$, OLR has approached the blackbody spectrum with a temperature equal to the cirrus temperature of 230 K. Hence, the effect of cirrus is to remove the layer emitting to space to a higher level and colder temperature. Outgoing radiation is most reduced in the atmospheric window between 8 and 12 μm , where, under cloud-free conditions, radiation emitted by the surface passes freely to space.

Contrary to the preceding treatment, real cloud has limited horizontal extent. A finite cloud scatters energy out its sides. This enables neighboring clouds to interact. Energy lost through the sides of a cloud exceeds SW absorption for aspect ratios as small as $\frac{1}{20}$ (Cox, 1981). Finite extent is therefore important for all but extensive stratiform. This dependence makes optical characteristics of cloud sensitive to the vertical distribution of microphysical properties. A cloud which has small droplets near its base is less reflective than one which has small droplets near its top because, in the former, much of the SW energy has already been scattered out the sides before the highly reflective droplets are encountered.

9.5 ROLES OF CLOUD AND AEROSOL IN CLIMATE

Cloud modifies the global energy balance by altering the absorption and scattering characteristics of the atmosphere. The largest variations of albedo and OLR in Fig. 1.34 are associated with cloud. Convection also supports large transfers of sensible and latent heat from the Earth's surface (Fig. 1.32). They represent another important heat source for the atmosphere. Radiative effects similar to cloud are introduced by aerosol, which controls cloud formation through microphysical processes (Sec. 9.2).

9.5.1 Involvement in the global energy budget

Influence of cloud cover

The role of cloud in global energetics may be interpreted as a forcing in the TOA radiative energy balance. For LW radiation, the large opacity of cloud increases the optical depth of the atmosphere. Enhancing the greenhouse effect (8.81), this favors increased surface temperature. Thus, cloud introduces warming in the LW energy budget. This warming varies with cloud-top temperature and height (Fig. 9.37). Cold cirrus also cools the lower stratosphere via exchange in the $9.6 \mu\text{m}$ band of ozone (Sec. 8.5.3). For SW radiation, the high reflectivity of cloud decreases the incoming solar flux. This favors reduced surface temperature (8.81). Thus, cloud introduces cooling in the SW energy budget. Unlike the LW effect, cloud albedo is insensitive to cloud height. Low cloud, like that in visible imagery (Fig. 1.29), is therefore almost as bright as high cloud. The zenith angle dependence of scattering (Fig. 9.33b) enhances albedo at high latitude. Cloud also introduces heating by absorbing SW radiation.

Cloud routinely covers about 50% of the Earth. It accounts for about half of the Earth's albedo (eg, for ~ 0.15). Marine stratocumulus (e.g., over the southeastern Atlantic in Fig. 1.29) is particularly important. Because of its extensive coverage, it reflects much of the incident SW radiation. However, because its cloud-top temperature does not differ substantially from that of the surface, marine stratocumulus does not appreciably modify OLR and hence the LW energy budget (cf. Fig. 1.28). In contrast, deep cumulus over Africa, South America, and the maritime continent has very cold cloud top temperature, as much as 100 K colder than the underlying surface. Consequently it sharply reduces OLR from that under cloud-free conditions.

The SW and LW effects of cloud are each variable. Owing to the zenith angle dependence of scattering (Fig. 9.33b) and to systematic variations in cloud cover and type, the most important component of cloud variability is diurnal. Figure 9.39 shows the daily variation of cloud properties over the southeastern Pacific, which is populated by marine stratocumulus, and over the Amazon basin, the site of deep continental cumulus. Marine stratocumulus (Fig. 9.39a) is present throughout the day. However, it undergoes a 50% variation in albedo. This resembles the variation of fractional cloud cover, which results from SW heating during daylight hours. The corresponding variation of OLR is small. Continental cumulus (Fig. 9.39b) undergoes a 25% variation of OLR. It resembles the daily life cycle of cumulonimbus tower and cirrostratus anvil. The corresponding variation of fractional cloud cover is small. Cirrostratus anvil is glaciated, markedly colder than the surface. It persists well into the night, several hours after convection has dissipated (Platt et al., 1984a).

A quantitative description of how cloud figures in the global energy budget is complicated by its dependence on microphysical properties and interactions with the

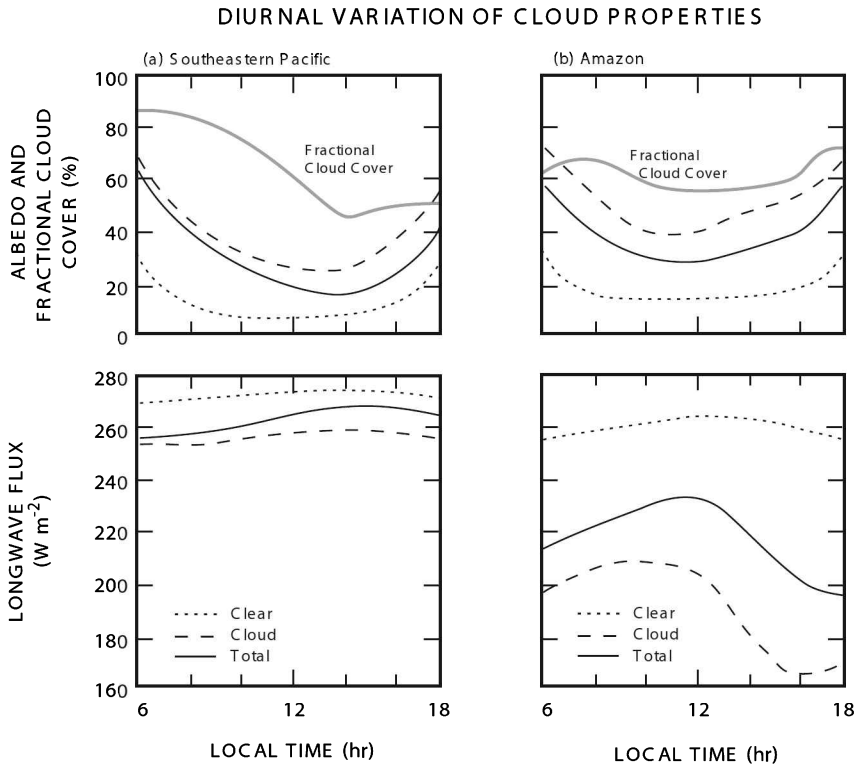


Figure 9.39 Top-of-the-atmosphere radiative properties, as functions of local time, over (a) the southeastern Pacific, which is populated by marine stratocumulus, and (b) the Amazon basin, which is populated by organized cumulonimbus and anvil cirrus. Adapted from Minnis and Harrison (1984).

surface. These complications are circumvented by comparing radiative fluxes at TOA under cloudy vs clear-sky conditions. Over a given region, the column-integrated radiative heating rate must equal the difference between the energy flux absorbed and that emitted to space

$$\begin{aligned}\dot{Q} &= \int_0^{\infty} \rho \dot{q} dz \\ &= (1 - \mathcal{A})F_s - F_{LW}(z = \infty),\end{aligned}\quad (9.48)$$

where \mathcal{A} is the albedo of the region and F_s and $F_{LW}(\infty)$ denote the downward SW and upward LW fluxes at TOA. The *cloud radiative forcing* is defined by

$$C = \dot{Q} - \dot{Q}^{CS}, \quad (9.49)$$

where \dot{Q}^{CS} is the heating rate under clear-sky conditions. It represents the net radiative influence of cloud on the column energy budget of the region. Incorporating (9.49) into (9.48) yields

$$C = (\mathcal{A}^{CS} - \mathcal{A})F_s + (F_{LW}^{CS} - F_{LW}). \quad (9.50)$$

From (9.50), the SW and LW components of cloud forcing are identified as

$$\begin{aligned} C_{SW} &= (\mathcal{A}^{CS} - \mathcal{A})F_s \\ &= F_{SW}^{CS} - F_{SW} \end{aligned} \quad (9.51.1)$$

$$C_{LW} = F_{LW}^{CS} - F_{LW}. \quad (9.51.2)$$

If the region has fractional cloud cover η , the outgoing SW and LW fluxes are given by

$$\begin{aligned} F_{SW} &= (1 - \eta)F_{SW}^{CS} + \eta F_{SW}^{OC} \\ F_{LW} &= (1 - \eta)F_{LW}^{CS} + \eta F_{LW}^{OC}, \end{aligned} \quad (9.52)$$

where the superscripts refer to clear-sky and overcast subregions. Then the SW and LW components of cloud radiative forcing may be written

$$\begin{aligned} C_{SW} &= \eta(F_{SW}^{CS} - F_{SW}^{OC}) \\ &= \eta F_s (\mathcal{A}^{CS} - \mathcal{A}^{OC}) \end{aligned} \quad (9.53.1)$$

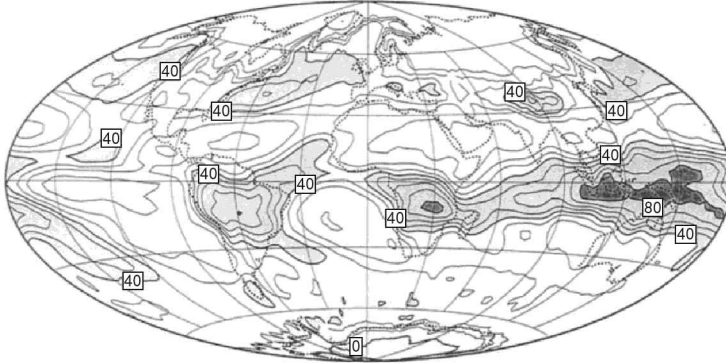
$$C_{LW} = \eta(F_{LW}^{CS} - F_{LW}^{OC}). \quad (9.53.2)$$

The components of cloud forcing (9.53) can be evaluated directly from broadband fluxes of outgoing LW and SW radiation that are measured by satellite. Figure 9.40 shows time-averaged distributions of C_{SW} , C_{LW} , and C . Longwave forcing (Fig. 9.40a) is large in centers of deep convection over tropical Africa, South America, and the maritime continent, where C_{LW} approaches 100 W m^{-2} (cf. Fig. 1.30b). Secondary maxima appear in the maritime ITCZ and in the North Pacific and North Atlantic storm tracks (Sec. 1.2.5). Shortwave forcing (Fig. 9.40b) is strong in the same regions, where $C_{SW} < -100 \text{ W m}^{-2}$. Negative SW forcing is also strong over extensive marine stratocumulus in the eastern oceans and over the Southern Ocean, coincident with the storm track of the Southern Hemisphere. Inside the centers of deep tropical convection, SW and LW cloud forcing nearly cancel. They leave small values of C throughout the tropics (Fig. 9.40c). Negative C_{SW} in the storm tracks and over marine stratocumulus then dominates positive C_{LW} , especially over the Southern Ocean. It prevails in the global-mean cloud forcing.

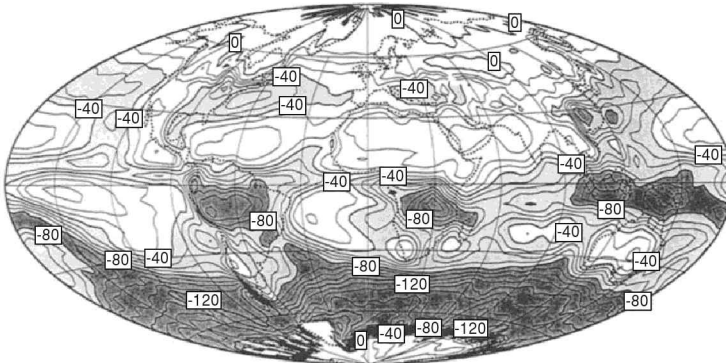
Globally averaged values of C_{LW} and C_{SW} are about 30 and -45 W m^{-2} , respectively. Net cloud forcing is then -15 W m^{-2} . It represents radiative cooling of the Earth-atmosphere system. This is four times as great as the additional warming of the Earth's surface that would be introduced by a doubling of CO_2 . Latent heat transfer to the atmosphere (Fig. 1.32) is 90 W m^{-2} . It is an order of magnitude greater. Consequently, the direct radiative effect of increased CO_2 would be overshadowed by even a small adjustment of convection (Sec. 8.7).

While circumventing many of the complications surrounding cloud behavior, cloud forcing is limited in several respects. Foremost is the fact that C gives only the column-integrated effect of cloud. It provides no information on the vertical distribution of that radiative heating and cooling. Shortwave cloud forcing represents cooling. It is concentrated near the Earth's surface, because the principal effect of increased albedo is to shield the ground from incident SW. Longwave cloud forcing represents warming. It is manifest in heating near the base of cloud and cooling near its top (Fig. 9.36b).

(a) Longwave Cloud Forcing



(b) Shortwave Cloud Forcing (SWCF)



(c) Net Radiative Cloud Forcing

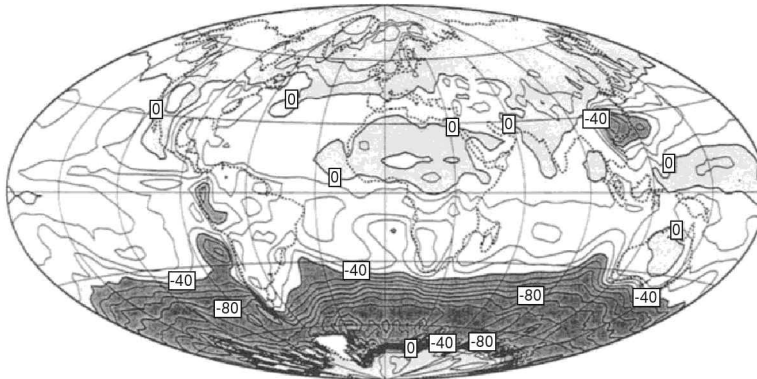


Figure 9.40 Cloud radiative forcing during northern winter derived from ERBE measurements on board the satellites ERBS and NOAA-9 for the (a) LW energy budget, (b) SW energy budget, and (c) net radiative energy budget. Courtesy of D. Hartmann (U. Washington).

That radiative forcing depends intrinsically on the vertical distribution of cloud. For instance, deep cumulonimbus and comparatively shallow cirrostratus can have identical cloud-top temperature, yielding the same LW forcing of the TOA energy budget. However, they have very different optical depths, producing very different vertical distributions of radiative heating. The strong correlation between water vapor and cloud cover introduces another source of uncertainty. Absorption by water vapor biases C_{LW} toward higher values (Hartmann and Doelling, 1991).

Cloud also introduces heating through the release of latent heat. Precipitation leads to a net transfer of heat from ocean to the atmosphere. Latent heating is particularly important for organized deep cumulus, which produces copious rainfall in the tropics. The column-integrated latent heating rate follows from the specific latent heat l and the precipitation rate \dot{P} as

$$\dot{Q} = l\rho_w\dot{P}. \quad (9.54)$$

\dot{Q} is a thermodynamic counterpart of TOA radiative fluxes in (9.48). According to (9.54), each mm day^{-1} of precipitation translates into $\sim 25 \text{ W m}^{-2}$ in the column energy budget of the atmosphere.

Figure 9.41a illustrates the geographical distribution of precipitation. Large precipitation is found inside the ITCZ, where \dot{P} exceeds 10 mm day^{-1} . There, latent heating introduces more than 250 W m^{-2} of warming into the column energy budget of the atmosphere, which is plotted in Fig. 9.41b. Acting in the opposite sense is radiative cooling over the eastern oceans. Characterized by cold SST and extensive marine stratocumulus, those regions introduce $50\text{--}150 \text{ W m}^{-2}$ of cooling into the column energy budget (cf. Fig. 1.29). The striking resemblance between latent heating and column-integrated heating in Fig. 9.41 makes the former a major energy source for the atmosphere. It also makes the atmosphere intrinsically sensitive to SST (Sec. 17.6).

The region of strong latent heating is the same one implied by cloud forcing to experience strong radiative heating and cooling. However, latent heating of the atmospheric column is three times greater than SW or LW forcing individually. Further, those components of radiative forcing experience strong cancellation inside convective centers, making latent heating an order of magnitude stronger than net radiative forcing by cloud. Latent heating is also distinguished by vertical structure. Whereas cloud radiative forcing is sharply concentrated in the vertical, latent heating inside deep cumulus influences much of the tropical troposphere. Latent heat release inside tropical convection is plotted in Fig. 9.42. It leads to a specific heating rate $\frac{\dot{q}}{c_p}$ of order 10 K day^{-1} . Through detrainment, that heat is transferred to the environment of deep cumulus (Sec. 7.7). Its deep distribution makes latent heat release a major source of energy for the tropical troposphere and the atmosphere as a whole (Fig. 9.41).

Influence of aerosol

Atmospheric aerosol introduces similar effects, albeit weaker. Like cloud, aerosol is an efficient scatterer of SW radiation. It alters optical properties following major volcanic eruptions (Fig. 9.5). Another source of aerosol is fire. Displayed on the cover is visible imagery from the GMS satellite (see color plate section). It reveals haze from Indonesian forest fires, which were supported by prolonged El Niño conditions (Sec. 17.6). Being entrained into twin tropical cyclones, such haze led to health restrictions and public closures. Anthropogenic sources of aerosol exert a similar influence. Near industrial centers in China, aerosol can become optically thick. There, reductions of SW flux at the surface of 10 W m^{-2} per decade have been reported (Norris and Wild, 2009). Yet,

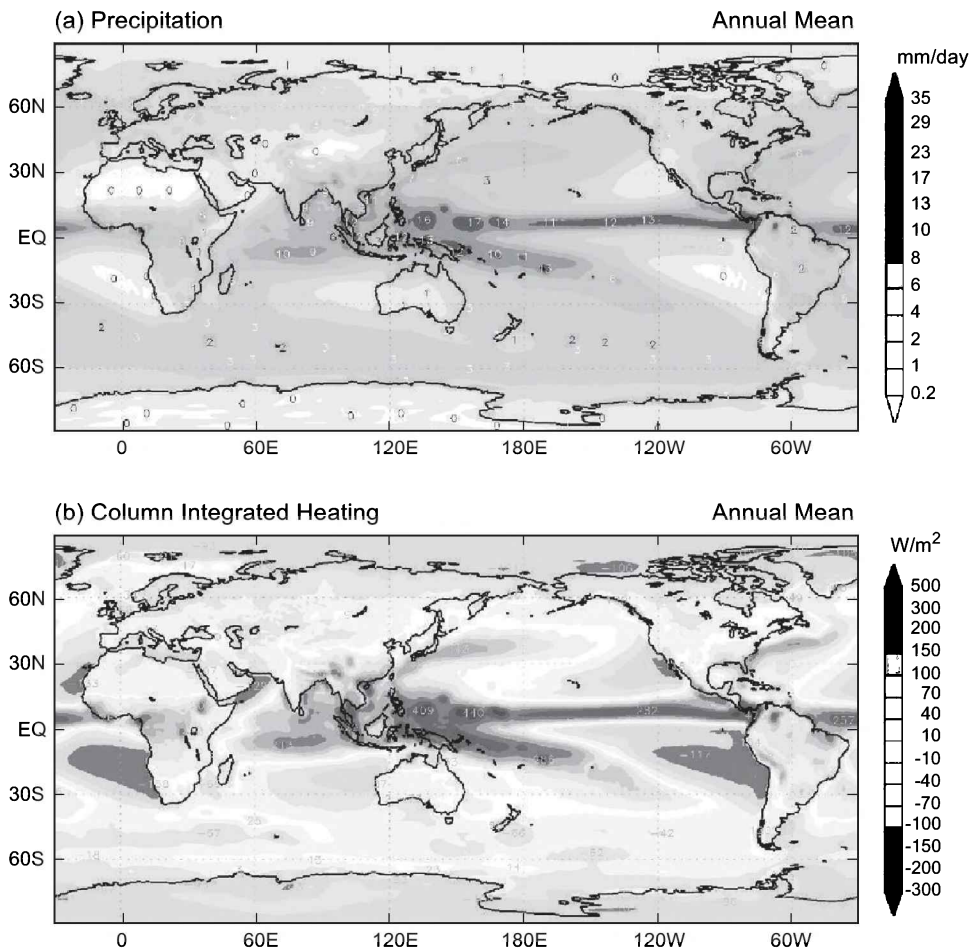


Figure 9.41 (a) Annual-mean precipitation rate (mm/day). (b) Column-integrated atmospheric heating ($W m^{-2}$). Both are concentrated inside the ITCZ, where precipitation and atmospheric heating exceed $10 mm/day$ and $250 W m^{-2}$, respectively. Support comes from the South Pacific Convergence Zone (SPCZ), the South Atlantic Convergence Zone (SACZ), and the North Atlantic and North Pacific storm tracks. *Source:* Japanese 25-yr Climatology, cooperative project of Japan Meteorological Agency (JMA) and Central Research Institute of Electric Power Industry (CRIEPI) (<http://ds.data.jma.go.jp>). See color plate section: Plate 7.

over Japan, the same period witnessed no significant change. Generally, such effects are confined to a neighborhood of aerosol source regions, by efficient scavenging processes associated with convection.

Depending on its composition and the underlying surface, aerosol can either enhance albedo or diminish it. Over dark ocean, aerosol increases albedo, introducing cooling into the SW energy budget. Conversely, over bright polar ice, arctic haze (involving anthropogenic pollutants), reduces albedo (Ackerman, 1988). Aerosol also

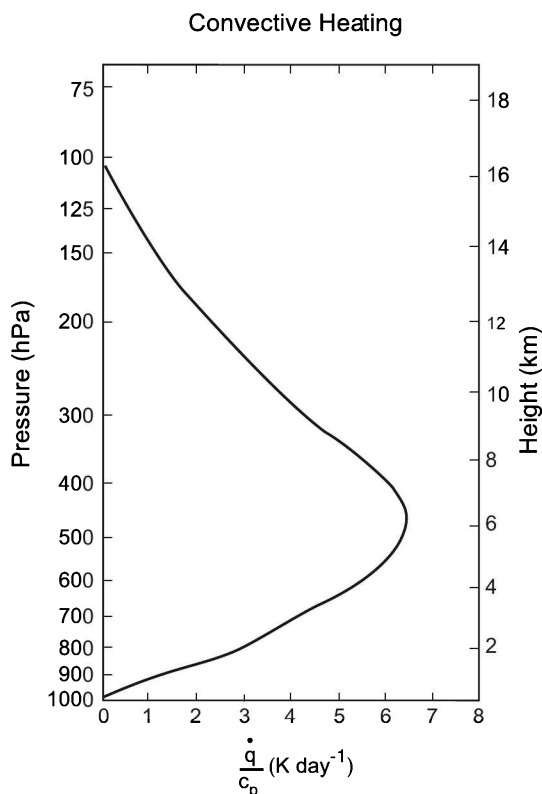


Figure 9.42 Vertical profile of the heating rate inside tropical convection, normalized by a precipitation rate of 10 mm day^{-1} . *Source:* Yanai et al. (1973).

contributes significantly to SW absorption inside the tropical troposphere (Fig. 8.26). To illustrate how these factors determine the net effect of aerosol, we consider a natural species: desert dust.

During disturbed periods, desert dust becomes optically thick. Once raised, it can be lifted inside vigorous desert convection as high as 500 hPa. Exposed to strong wind, it is then carried hundreds, sometimes thousands, of kilometers from its source. Figure 9.43 displays visible imagery during September 2009, when the eastern seaboard of Australia was blanketed by red dust that originated a thousand kilometers to the west. At some sites, the Earth's surface is entirely masked; the dust cloud is optically thick. It streams well off the continent, into the Pacific. Notice the change of albedo relative to neighboring ocean. Conditions inside the layer that is optically thick are displayed in Fig. 9.44, compared against the same field of view under normal conditions. Inside the layer that is optically thick, scattering is isotropic. Visibility is limited to a few tens of meters.

The influence of dust on the energy budget depends on its composition in relation to optical properties of the underlying surface. Dust absorbs SW radiation, reducing albedo over bright surface, like shallow stratus. It then introduces warming into the SW energy budget of the column (9.48). However, over dark surface like ocean, dust increases the albedo (Fig. 9.43). It then introduces cooling into the SW energy budget.

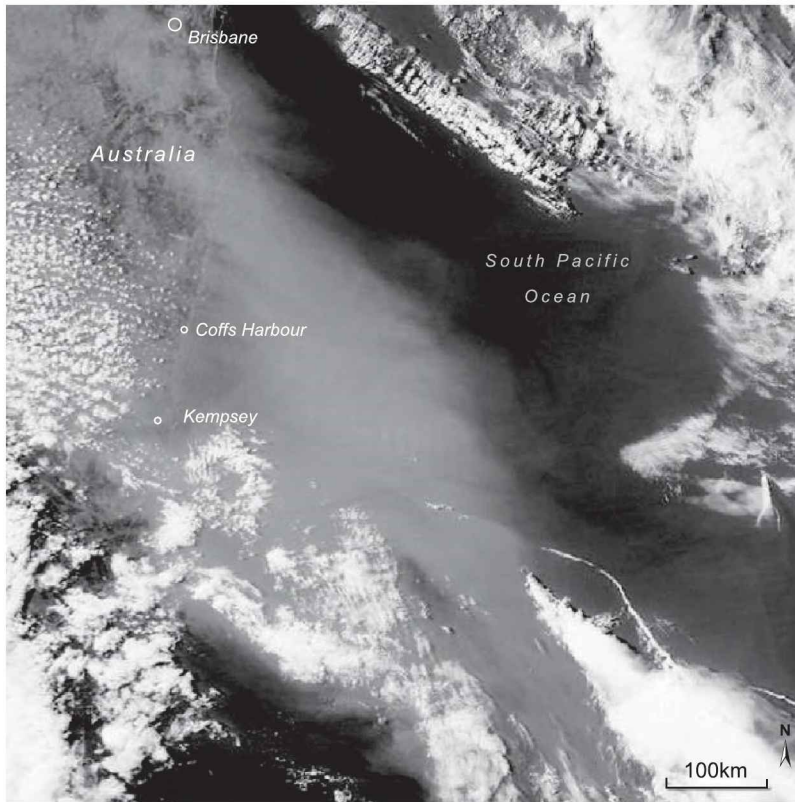


Figure 9.43 Visible imagery during September 2009, when the eastern seaboard of Australia was blanketed by red dust that originated 1000 km to the west. Notice the change of albedo relative to neighboring ocean. See color plate section: Plate 8.

The LW energy budget is influenced by desert aerosol in a manner similar to cloud. When convected to great height, desert aerosol can reduce the effective emission temperature, with commensurate warming in the column energy budget.⁸

The energy budget of desert aerosol is dominated by SW effects. Over surfaces with appreciable albedo, like the Sahara, those SW effects introduce net warming of the column. The same holds for the atmosphere alone. The vertical distribution of radiative heating, however, is determined by both LW and SW components. Figure 9.45 plots the net heating rate for Saharan dust that is distributed between the Earth's surface and 500 hPa. Over cloud-free ocean (Fig. 9.45a), absorption of upwelling LW radiation introduces significant heating near the surface, where $\frac{\dot{q}}{c_p} > 2 \text{ K day}^{-1}$. For optical depths greater than 1, SW absorption introduces a second heating maximum near the middle of the dust layer. Over cloud-free desert (Fig. 9.45b), greater surface albedo backscatters much of the transmitted SW radiation, which is then absorbed inside

⁸ The reduction of effective emission temperature can exceed that of cloud at the same height because, unlike cloud, temperature at the top of the dust layer decreases with height dry adiabatically.



Figure 9.44 Field of view across Sydney Harbor (a) under normal conditions and (b) during the dust storm of September 2009. With permission, copyright 2010 Fairfax Digital. See color plate section: Plate 9.

Atmospheric Heating by Saharan Dust

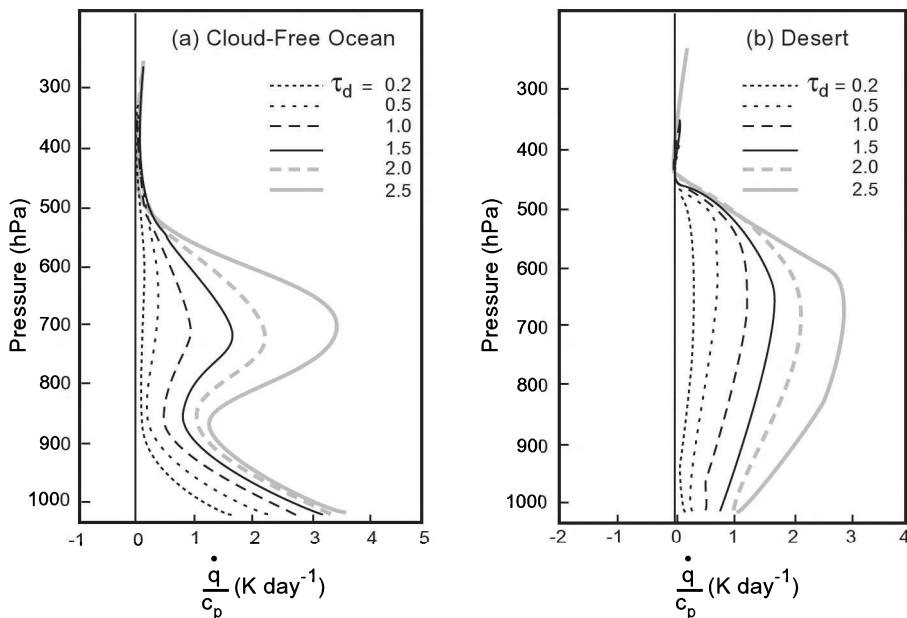


Figure 9.45 Vertical profiles of heating rate for Saharan dust that is distributed between 1000 hPa and 500 hPa and has optical depths τ_d over (a) cloud-free ocean and (b) desert. Adapted from Carlson and Benjamin (1982). Copyright by Spectrum Press (A. Deepak Publishing).

the dust layer. Net heating is thus dominated by SW absorption. It leads to $\frac{d}{c_p}$ of order $1-2 \text{ K day}^{-1}$ across much of the dust layer. Because it heats the atmosphere and cools the underlying surface, the net effect in both cases is to stabilize the stratification.

Beyond its direct radiative effect, aerosol influences cloud formation. This role of aerosol is potentially more significant than its direct radiative effect. Aerosol underpins the microphysical properties of cloud, like condensation, droplet size distribution, and precipitation (Rosenfeld, 2006). As these features shape the evolution of cloud, indirectly, aerosol influences macrophysical properties such as cloud growth and latent heating. An outbreak of desert dust alters the coverage and type of cloud, favoring stratiform. Through such dependence, a change in the concentration, size, or composition of atmospheric aerosol can introduce changes in cloud cover and in the radiative energy budget.

9.5.2 Involvement in chemical processes

Cloud also supports important interactions with chemical species. Precipitation is the primary removal mechanism for atmospheric pollutants. By serving as CCN, aerosol particles are incorporated into cloud droplets and precipitated to the Earth's surface. This process scavenges air of most nuclei activated and others absorbed through collision with cloud droplets. The efficiency of washout accounts for the short residence time of tropospheric aerosol (of order a day). Washout also operates on gaseous pollutants that are water soluble.

Aerosol particles support heterogeneous chemistry, which requires the presence of multiple phases. Cloud particles that comprise PSC are involved in the formation of the Antarctic ozone hole (Chap. 18). PSC also introduces an important transport mechanism, which can likewise influence ozone depletion. If they attain sufficient size, PSC particles undergo sedimentation - like cirrus ice particles. Water vapor and nitrogen are then removed from the lower stratosphere.

Chemical reactions similar to those occurring inside PSC take place on natural aerosol. The concentration of natural aerosol increases sharply following major volcanic eruptions (Fig. 9.1). Enhanced aerosol surface area then becomes comparable to that normally observed over the Antarctic in connection with PSC. This mechanism may have figured in diminished ozone that was observed following the eruption of El Chichon (Hoffman and Solomon, 1989). More generally, however, ozone, even over the Antarctic, undergoes large changes from one year to the next. They are introduced through changes of the circulation (Chap. 18). Influencing transport and temperature, interpreting such changes requires an understanding of air motion.

SUGGESTED REFERENCES

Atmospheric aerosol is discussed in the monograph *Atmospheric Aerosols: Their Formation, Optical Properties, and Effects* (1982), edited by Deepak.

An overview of cloud microphysics is presented in *Atmospheric Science: An Introductory Survey* (2006) by Wallace and Hobbs. Advanced treatments can be found in *Microphysical Processes in Clouds* (1993) by Young and in *Microphysics of Clouds and Precipitation* (1978) by Pruppacher and Klett.

International Cloud Atlas (WMO, 1969) presents a complete classification of cloud.

Cloud Dynamics (1993) by Houze is a detailed treatment of the subject. An advanced treatment, including laboratory simulations, is presented in *Environmental Aerodynamics* (1978) by Scorer. *Classical Electrodynamics* (1975) by Jackson develops Rayleigh scattering. A comprehensive treatment of cloud radiative processes is provided in *Radiation and Cloud Processes in the Atmosphere* (1990) by Liou.

Hartmann et al. (1986) contains an overview of how cloud and surface influence the Earth's energy budget.

PROBLEMS

1. In terms of the size distribution $n(a)$, derive an expression for the size spectrum of (a) aerosol surface area $S(a)$ and (b) aerosol mass $M(a)$, if particles have uniform density ρ .
2. A population of cloud droplets has the log-normal size spectrum

$$\frac{dn}{d(\log a)} = \frac{n_0}{\sigma\sqrt{2\pi}} e^{-\frac{\ln^2\left(\frac{a}{a_m}\right)}{2\sigma^2}},$$

where $n_0 = 300 \text{ cm}^{-3}$, $a_m = 5 \text{ }\mu\text{m}$ is the median droplet radius, and $\sigma = 4 \text{ }\mu\text{m}$ is the rms deviation of $\ln(a)$. (a) Plot the size spectrum $\frac{dn}{d(\log a)}$, the surface area spectrum $\frac{dS}{d(\log a)}$, and the mass spectrum $\frac{dM}{d(\log a)}$. (b) Contrast the distributions of those properties over droplet radius a .

3. Provide an expression for the potential barrier that must be overcome for a water droplet to be activated.
4. (a) Determine the critical radius at 0.2% supersaturation for a droplet of pure water. (b) How large must a droplet containing 10^{-16} g of sodium chloride become to grow spontaneously?
5. A cumulus cloud forms inside a moist thermal. The base of the thermal is at 1000 hPa, where $T_0 = 30^\circ\text{C}$ and $r_0 = 10 \text{ g kg}^{-1}$. (a) If, at all elevations, the thermal is 2°C cooler near its periphery than in its core (other factors being equivalent), determine the height difference between the cloud base at its center versus at its periphery. (b) How would mixing of moisture with the environment modify this result?
6. The flow about a falling sphere of radius a and velocity w is *laminar* if the dimensionless Reynolds number

$$Re = \frac{\rho a w}{\mu}$$

is less than 5000 (Chap. 13), where $\mu = 1.7 \cdot 10^{-5} \text{ kg m}^{-1}\text{s}^{-1}$ is the coefficient of viscosity for air. Under these circumstances, the sphere experiences a viscous force

$$D = -6\pi\mu a w$$

known as *Stokes drag*. (a) Calculate the terminal velocity w_t of a spherical cloud droplet of radius a . (b) For what range of a does the preceding result apply if the flow becomes turbulent for $Re > 5000$? (c) If the flow about them remains laminar, how large must droplets grow before they can no longer be supported by an updraft of 1 m s^{-1} ?

7. Consider a population of cloud droplets having size distribution $n(a)$ under the conditions of Prob. 9.6. (a) Derive an expression for the specific drag force exerted on air by the falling droplet population. (b) What is required to maintain an updraft in the presence of those droplets?
8. Apply the results of Prob. 9.6 to estimate the time for volcanic dust particles of density $2.6 \times 10^3 \text{ kg m}^{-3}$ to reach the Earth's surface from 50 hPa if the tropopause is at 200 hPa and the particles are spherical with a mean radius of (a) $1.0 \text{ }\mu\text{m}$, (b) $0.1 \text{ }\mu\text{m}$.
9. When the ground is snow covered, the sky appears brighter. Why?
10. Rayleigh scattering intensifies inversely with wavelength. (a) Why is the sky not violet? (b) Use the Planck spectrum for an equivalent blackbody temperature of 6000 K to show that scattered SW radiation maximizes at wavelengths of blue light.
11. Calculate the atmospheric albedo due to Rayleigh scattering for overhead sun and (a) $\lambda = 0.7 \text{ }\mu\text{m}$, (b) $\lambda = 0.5 \text{ }\mu\text{m}$, (c) $\lambda = 0.3 \text{ }\mu\text{m}$. The refractive index of air at 1000 hPa and 273°K is approximated by

$$(m - 1)10^6 = 6.4328 \times 10^1 + \frac{2.94981 \times 10^4}{146 - \lambda^{-2}} + \frac{2.554 \times 10^2}{41 - \lambda^{-2}},$$

with λ in μm and $m - 1$ proportional to density.

12. Following the eruption of Krakatoa, sunsets were modified for several years. How small must particles have been?
13. (a) Use the mean latent heat flux to the atmosphere: 90 W m^{-2} (Fig. 1.32), to estimate the global-mean precipitation rate. (b) Use the result in (a) to estimate the fraction of the water vapor column in the tropics that is processed by convection during one day. Reconcile this fraction with the coverage by and time scale of deep convection.
14. Use the mean annual precipitation in Fig. 9.41 and a characteristic tropopause height of 16 km to estimate the latent heating rate inside the ITCZ.
15. Estimate the entrainment height (Sec. 7.4.2) for a thermal of diameter (a) 100 m, (b) 1 km.
16. A stack of lenticular forms from moisture anomalies upstream that are sheared and advected through the mountain wave by the prevailing flow. Describe their evolution.
17. The milky appearance of the sky, often observed from the surface, disappears at an altitude of a kilometer or two. Why?
18. Nimbus cloud generally appears darker than other cloud. Why?
19. Use the scattered intensity (9.19) to obtain the total scattered power (9.20) for Rayleigh scattering.
20. Transmission of diffuse SW radiation through a scattering layer is described in terms of the phase function $P(\cos\Theta)$, where Θ is the 3-dimensional scattering angle. Express $\cos\Theta$ in terms of the zenith angles θ and θ' and the azimuthal angles ϕ and ϕ' of scattered and incident radiation, respectively.
21. Scattering of microwave radiation by cloud droplets and precipitation enables cloud properties to be measured by radar. (a) Verify that Rayleigh scattering is a valid description of such behavior. (b) The backscattering coefficient (8.19), which measures the reflected power per unit length of scattering medium, follows from

the scattering cross section as

$$\beta_s(\pi) = n\hat{\sigma}_s P(\pi),$$

where n is the particle number density and $P(\pi)$ is the phase function for backward scattering. Derive an expression for $\beta_s(\pi)$ in terms of the radius a of scatterers and the fractional volume $\eta = n \cdot \frac{4}{3}\pi a^3$ occupied by them to show that, for fixed η (eg, fixed liquid water content), the reflectivity is proportional to a^6 .

22. Show that the blackbody function $B^*[T]$ is a particular solution of (9.46).
23. For the cloud droplet population in Prob. 9.2, compute (a) the scattering-equivalent mean droplet radius, (b) the liquid water path for a cloud 4 km deep, (c) the cloud's optical depth, (d) the cloud's absorptivity for $\lambda = 0.5 \mu\text{m}$ and a solar zenith angle of 30° .
24. Calculate the optical depth posed to SW radiation by a 5-km cumulus congestus (an organization of cumulus), which has an average liquid water content of 0.66 g m^{-3} and a scattering-equivalent mean droplet radius of $12 \mu\text{m}$.
25. A homogeneous cloud layer of optical depth τ_c is illuminated by the solar flux F_s at a zenith angle $\theta_s = \cos^{-1}\mu_s$. If the surface is gray with absorptivity a , (a) derive expressions for the upwelling and downwelling diffuse fluxes inside the scattering layer, (b) plot the albedo and transmissivity, as functions of a , for overhead sun and $\tau_c = 20$.
26. Consider the radiative-convective equilibrium for the cloudy atmosphere in Fig. 9.36. (a) Explain why the surface temperature in Fig. 9.37 increases linearly with cloud height once the cloud is elevated sufficiently. (b) Nearly identical results follow for other cloud optical depths. Why?
27. Satellite observations reveal the following properties for a region of maritime convection in the tropics: fractional cloud cover of 0.5, average clear-sky and overcast albedos of 0.2 and 0.8, respectively, and average clear-sky and overcast LW fluxes of 350 W m^{-2} and 100 W m^{-2} , respectively. Calculate the average SW, LW, and net cloud forcing for the region.
28. Use the directionality of scattering to assess the change in overall SW transmission accompanying the reduction of direct transmission following the eruption of El Chichon in Fig. 9.5.
29. Satellite measurements of SW radiation enable the albedo of cloud to be determined. From it, cloud optical depth can be inferred. (a) Plot cloud albedo as a function of optical depth for a solar zenith angle of 30° , presuming a homogeneous scattering layer and a black underlying surface. (b) Discuss the sensitivity of cloud albedo to changes in optical depth for $\tau_c > 50$. What implications does this have for inferring optical depth?
30. A satellite measuring backscattered SW radiation and LW radiation emitted at $11 \mu\text{m}$ observes a tropical cloud with an albedo of 0.9 and a brightness temperature of 220 K. If the solar zenith angle is 30° and a scattering-equivalent radius of $5 \mu\text{m}$ applies, use the result of Prob. 9.29 to estimate (a) the liquid water path of the cloud, (b) the cloud's average liquid water content if it is 1 km thick.

Atmospheric motion

Under radiative equilibrium, the First Law reduces to a balance between radiative transfers of energy. This simple energy balance determines the radiative-equilibrium thermal structure. It is valid for a resting atmosphere. However, radiative equilibrium breaks down for an atmosphere in motion. Heat is then also transferred mechanically, by air motion. The radiative-equilibrium thermal structure (Fig. 8.21) differs conspicuously from the observed global-mean temperature (Fig. 1.2). Under radiative equilibrium, temperature is much too warm at the surface. It decreases upward too steeply. The discrepancy from observed temperature points to the importance of mechanical heat transfer. Under radiative-convective equilibrium, mechanical heat transfer is accounted for – by convection. It reconciles the equilibrium thermal structure with that observed, but only vertical structure that is characteristic of the global-mean.

Horizontal thermal structure differs significantly from that observed (Fig. 10.1). Under radiative-convective equilibrium (Fig. 10.1a), temperature decreases poleward at all latitudes, from ~ 315 K over the equator to colder than 180 K over the winter pole. Observed zonal-mean temperature (Fig. 10.1b) decreases poleward more gradually. It is some 20 K cooler over the equator and 40 K warmer over the winter pole. By comparison, radiative-convective equilibrium produces a meridional temperature gradient that is too steep. It accounts for mechanical transfer of heat, but only in the vertical.

At observed temperatures, individual latitudes are not in radiative equilibrium, not even in radiative-convective equilibrium. According to the distribution of net radiation (Fig. 1.34c), low latitudes experience radiative heating. Conversely, middle and high latitudes experience radiative cooling. To maintain thermal equilibrium locally, those radiative imbalances must be compensated by a poleward transfer of heat: from the tropics, where the radiative energy budget is in surplus, to middle and high

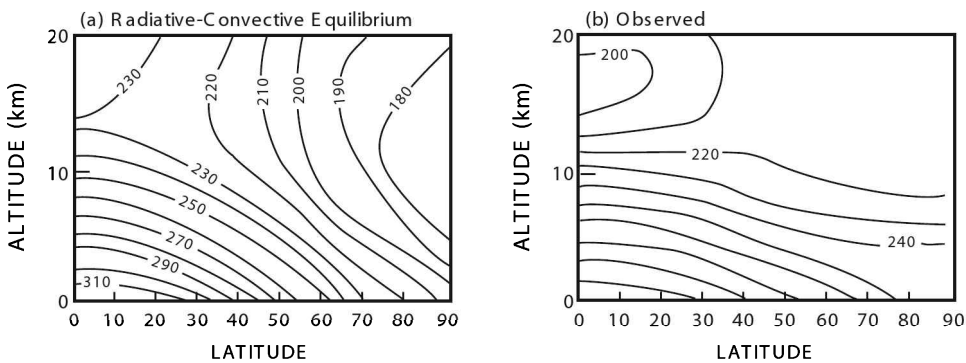


Figure 10.1 Zonal-mean temperature as a function of latitude and height (a) under radiative-convective equilibrium and (b) observed during northern winter. Without horizontal heat transfer, radiative-convective equilibrium establishes a meridional temperature gradient that is much steeper than observed. *Sources:* Liou (1990) and Fleming et al. (1988).

latitudes, where it is in deficit. This horizontal heat transfer is accomplished by the general circulation of the atmosphere and ocean, most of it by the atmospheric circulation. Understanding how observed thermal structure is maintained therefore requires an understanding of atmospheric motion and horizontal heat transfer that it accomplishes.

10.1 DESCRIPTION OF ATMOSPHERIC MOTION

Until now, the development has focused on an individual air parcel, a material element of infinitesimal dimension. Thermodynamic and hydrostatic influences acting on that discrete system must now be complemented by dynamical considerations that control its movement. Together, these properties constitute the individual or *Lagrangian description* of fluid motion. The latter represents atmospheric behavior in terms of the collective properties of material elements that comprise the atmosphere. Because the basic laws of physics apply to a discrete bounded system, they are developed most simply within the Lagrangian framework. Despite this conceptual advantage, in practice, the Lagrangian description of atmospheric behavior can be cumbersome, especially in numerical applications. It requires representing not only the thermal, mechanical, and chemical histories of individual air parcels, but also tracking their positions - not to mention their distortion as they move through the circulation.

Like any fluid system, the atmosphere is a continuum. It is therefore composed of infinitely many such discrete systems. In principle, all of them must be represented in the Lagrangian description of atmospheric behavior. For this reason, it is more convenient to describe atmospheric behavior in terms of field variables, which represent the distributions of properties at a particular instant. The distribution of temperature $T(\mathbf{x}, t)$, where $\mathbf{x} = (x, y, z)$ denotes 3-dimensional position, is a scalar field variable. So is the distribution of mixing ratio $r(\mathbf{x}, t)$. Likewise for the component of velocity in the i th coordinate direction $v_i(\mathbf{x}, t)$. Collecting the scalar components of velocity in the three coordinate directions gives the 3-dimensional velocity field $\mathbf{v}(\mathbf{x}, t) = v_i(\mathbf{x}, t)$; $i = 1, 2, 3$. Velocity is a vector field variable. At each location, it has both magnitude and direction.

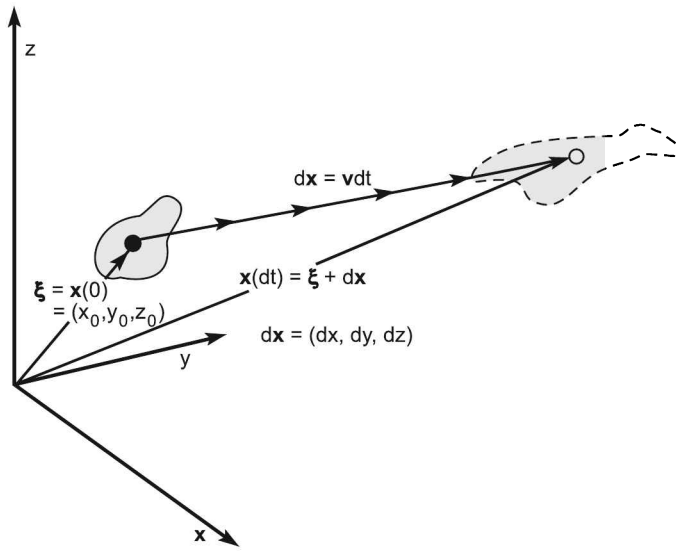


Figure 10.2 Positions \mathbf{x} of a material element initially and after an interval dt . The element's initial position $\mathbf{x}_0 = (x_0, y_0, z_0)$ defines its *material coordinate* ξ . During the interval dt , the element traces out an increment of *parcel trajectory* $d\mathbf{x} = \mathbf{v} dt$, where \mathbf{v} is the element's instantaneous velocity.

Like thermodynamic and chemical variables, the velocity field $\mathbf{v}(\mathbf{x}, t)$ is a property of the fluid system. It describes the circulation. Because air motion transfers heat and chemical constituents, $\mathbf{v}(\mathbf{x}, t)$ is coupled to other field properties. Collectively, the distributions of such properties constitute the *field* or *Eulerian description* of fluid motion. The latter is governed by the equations of continuum mechanics.

The Eulerian description simplifies the representation of atmospheric behavior. However, physical laws governing atmospheric behavior, like the first and second laws of thermodynamics and Newton's laws of motion, apply directly to a fixed collection of matter. For this reason, the equations governing atmospheric behavior are developed most intuitively in the Lagrangian framework. In the Eulerian framework, the field property at a specified location involves different material elements at different times. Despite this complication, the Eulerian description is related to the Lagrangian description by a fundamental kinematic constraint: The field property at a given location and time must equal the property possessed by the material element which occupies that position at that time.

10.2 KINEMATICS OF FLUID MOTION

A material element (Fig. 10.2) can be identified by its initial position

$$\begin{aligned}\xi &= \mathbf{x}_0 \\ &= (x_0, y_0, z_0),\end{aligned}\tag{10.1.1}$$

which is referred to as the *material coordinate* of that element. The locus of points

$$\mathbf{x}(t) = (x(t), y(t), z(t))\tag{10.1.2}$$

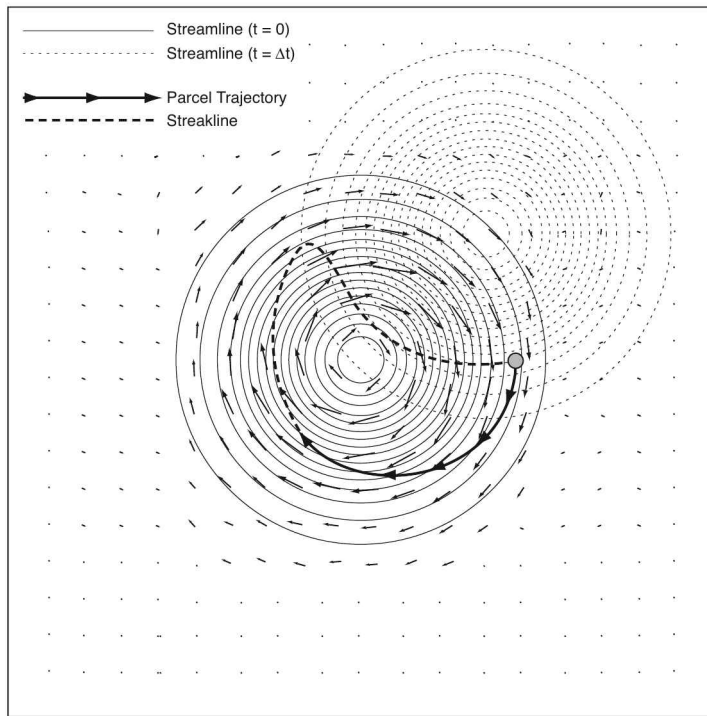


Figure 10.3 Motion and streamlines initially (solid) and streamlines after an interval Δt (dashed) for a vortex translating uniformly. Superposed is (1) the *parcel trajectory* at time Δt (bold solid) of a material element that initially coincides with the stippled circle and (2) the *streakline* at time Δt (bold dashed) that originates from the same position.

traced out by the element defines a *material* or *parcel trajectory*. It is uniquely determined by the material coordinate $\boldsymbol{\xi} = \boldsymbol{x}_0$ and the velocity field $\boldsymbol{v}(\boldsymbol{x}, t)$. With $\boldsymbol{\xi}$ held fixed, the vector $d\boldsymbol{x} = (dx, dy, dz)$ describes the incremental displacement of the material element during the interval dt . The element's velocity is then

$$\boldsymbol{v} = \frac{d\boldsymbol{x}}{dt}, \quad (10.2.1)$$

with components

$$\begin{aligned} u &= \frac{dx}{dt} \\ v &= \frac{dy}{dt} \\ w &= \frac{dz}{dt}. \end{aligned} \quad (10.2.2)$$

Because it is evaluated for fixed $\boldsymbol{\xi}$, the time derivative in (10.2) corresponds to an individual material element. It therefore represents the Lagrangian time rate of change, the rate of change on which the development has focused until now. By the kinematic constraint that relates Eulerian and Lagrangian descriptions, the velocity in (10.2) must equal the field value $\boldsymbol{v}(\boldsymbol{x}, t)$ at the material element's position $\boldsymbol{x}(t)$ at time t .

The velocity field defines a family of *streamlines* (Fig. 10.3), which, at any instant, are everywhere tangential to $\mathbf{v}(\mathbf{x}, t)$. If local velocities do not change with time, the motion field is *steady*. Parcel trajectories then coincide with streamlines. This feature of the motion follows from (10.2) because a parcel is then always displaced along the same streamline, which remains fixed. Under unsteady conditions, streamlines move to new positions. Parcel trajectories then no longer coincide with streamlines. Initially tangential to one streamline, the velocity of a parcel will displace it to a different streamline, with which its velocity is again temporarily tangential.

The motion is also characterized by a *streakline*. The latter is the locus of points traced out by a dye that is released into the flow at a particular location. As illustrated in Fig. 10.3, a streakline represents the positions at a given time of contiguous material elements that have passed through the location where the tracer is released. Like parcel trajectories, streaklines coincide with streamlines only if the velocity field is steady.

In addition to translation, a material element can undergo rotation and deformation as it moves through the circulation. These effects are embodied in the velocity gradient

$$\nabla \mathbf{v} = \frac{\partial v_i}{\partial x_j}, \quad i, j = 1, 2, 3 \quad (10.3)$$

which is a 2-dimensional tensor with the subscripts i and j referring to three Cartesian coordinate directions. The local velocity gradient $\nabla \mathbf{v}(\mathbf{x})$ determines the relative motion between material coordinates. It therefore describes the distortion experienced by the material element located at \mathbf{x} . Two material coordinates that are separated initially by dx_j experience relative motion

$$\begin{aligned} dv_i &= \frac{\partial v_i}{\partial x_j} dx_j, \\ &= \frac{\partial v_i}{\partial x_1} dx_1 + \frac{\partial v_i}{\partial x_2} dx_2 + \frac{\partial v_i}{\partial x_3} dx_3, \end{aligned} \quad (10.4)$$

where repeated indices in the same term imply summation and the vector product – the so-called *Einstein notation*. How the material element evolves through relative motion is elucidated by separating the velocity gradient into symmetric and anti-symmetric components:

$$\frac{\partial v_i}{\partial x_j} = e_{ij} + \omega_{ij}, \quad (10.5.1)$$

where

$$e_{ij} = \frac{1}{2} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \quad (10.5.2)$$

$$\omega_{ij} = \frac{1}{2} \left(\frac{\partial v_i}{\partial x_j} - \frac{\partial v_j}{\partial x_i} \right). \quad (10.5.3)$$

The symmetric component $e(\mathbf{x}, t)$ of the velocity gradient defines the *rate of strain* or *deformation tensor* acting on the material element at \mathbf{x} . Diagonal components of e describe variations of velocity “in the direction of motion.” They characterize the rate of longitudinal deformation, or “stretching,” in the three coordinate directions. For example, e_{11} describes the rate of elongation of a material segment

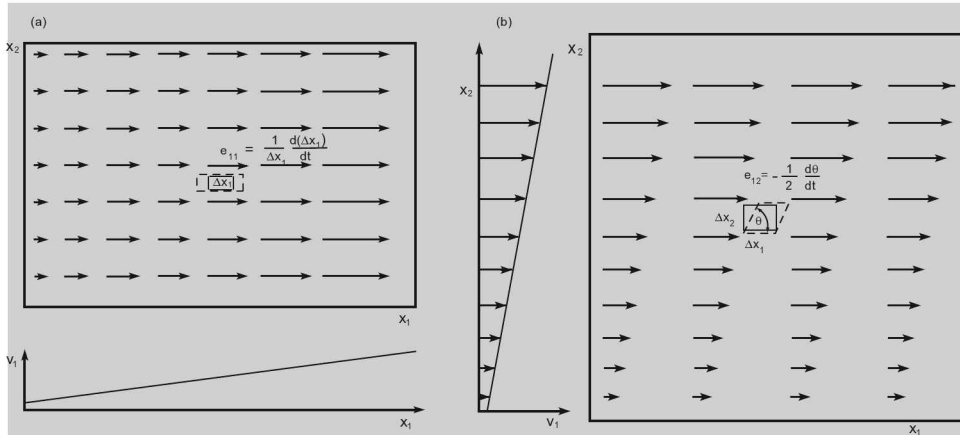


Figure 10.4 (a) Stretching of a material element in a longitudinally varying flow. The element’s rate of elongation is reflected in the *normal strain rate* e_{11} . (b) Shearing of a material element in a transverse-varying flow. The rate at which the material angle θ decreases is reflected in the *shear strain rate* e_{12} .

aligned with coordinate 1 (Fig. 10.4a). Collecting all three components of stretching gives the *trace* of \mathbf{e} , which describes the divergence of the velocity field

$$\frac{\partial v_i}{\partial x_i} = \nabla \cdot \mathbf{v}. \tag{10.6}$$

It represents the rate of increase in volume or *dilatation* of the material element. Under pure shear, the field $\mathbf{v}(\mathbf{x}, t)$ varies only in directions transverse to the motion. Diagonal elements of $\mathbf{e}(\mathbf{x}, t)$ then vanish, so $\nabla \cdot \mathbf{v} = 0$: The material element may experience distortion, but no change of volume.

Off-diagonal components of $\mathbf{e}(\mathbf{x}, t)$ describe variations of velocity “orthogonal to the motion.” They characterize the rate of transverse deformation, or “shear,” experienced by the material element at \mathbf{x} . As pictured in Fig. 10.4b, $2e_{12}$ represents the rate of decrease of the angle between two material segments aligned initially with coordinates 1 and 2. In the case of rectilinear motion that varies only longitudinally, off-diagonal components of $\mathbf{e}(\mathbf{x}, t)$ vanish. The material element at \mathbf{x} may then experience stretching, but no shear.

The antisymmetric component of $\nabla \mathbf{v}(\mathbf{x}, t)$ also describes relative motion, but not deformation. Because it contains only three independent components, the tensor $\boldsymbol{\omega}$ defines a vector

$$\hat{\boldsymbol{\omega}} = \begin{bmatrix} \omega_{23} \\ -\omega_{13} \\ \omega_{12} \end{bmatrix}. \tag{10.7}$$

The vector $\hat{\boldsymbol{\omega}}(\mathbf{x}, t)$ is related to the curl of the motion field or *vorticity*

$$2\hat{\boldsymbol{\omega}} = \nabla \times \mathbf{v}, \tag{10.8}$$

which represents twice the local angular velocity. Thus, $\hat{\boldsymbol{\omega}}(\mathbf{x}, t)$ represents the angular velocity of the material element at position \mathbf{x} . If symmetric components of

$\nabla \mathbf{v}(\mathbf{x}, t)$ vanish, so does the local deformation tensor $\mathbf{e}(\mathbf{x}, t)$. The material element's motion then reduces to translation plus rigid body rotation.

10.3 THE MATERIAL DERIVATIVE

To transform to the Eulerian description, the Lagrangian derivative that appears in conservation laws must be expressed in terms of field properties. Consider a field variable $\psi = \psi(x, y, z, t)$. The incremental change of property ψ is described by the total differential

$$\begin{aligned} d\psi &= \frac{\partial \psi}{\partial t} dt + \frac{\partial \psi}{\partial x} dx + \frac{\partial \psi}{\partial y} dy + \frac{\partial \psi}{\partial z} dz \\ &= \frac{\partial \psi}{\partial t} dt + \nabla \psi \cdot d\mathbf{x}, \end{aligned} \quad (10.9)$$

where dx, dy, dz , and dt are suitably defined increments in space and time. Let $d\mathbf{x} = (dx, dy, dz)$ and dt in (10.9) denote increments of space and time with the material coordinate $\boldsymbol{\xi}$ fixed. Then $d\mathbf{x}$ represents the displacement of the material element $\boldsymbol{\xi}$ during the time increment dt . It describes the increment of parcel trajectory shown in Fig. 10.2. With position and time related in this manner, the total differential describes the incremental change of property ψ observed "in a frame moving with the material element." Differentiating with respect to time (with $\boldsymbol{\xi}$ fixed) gives the time rate of change of ψ "following a material element":

$$\begin{aligned} \frac{d\psi}{dt} &= \frac{\partial \psi}{\partial t} + u \frac{\partial \psi}{\partial x} + v \frac{\partial \psi}{\partial y} + w \frac{\partial \psi}{\partial z} \\ &= \frac{\partial \psi}{\partial t} + \mathbf{v} \cdot \nabla \psi. \end{aligned} \quad (10.10)$$

It defines the *Lagrangian* or *material derivative* of the field variable $\psi(\mathbf{x}, t)$.

According to (10.10), the material derivative includes two contributions. The first, $\frac{\partial \psi}{\partial t}$, is the *Eulerian* or *local derivative*. It represents the rate of change of the material property ψ introduced by temporal changes in the field variable at the position \mathbf{x} where the material element is located. The second contribution, $\mathbf{v} \cdot \nabla \psi$, is the *advective contribution*. It represents the change of the material property ψ due to motion of the material element to positions of different field values. Even if the field is steady, namely if $\psi(\mathbf{x}, t) = \psi(\mathbf{x})$ so that local values do not change with time, the property of a material element does change if that element moves across contours of $\psi(\mathbf{x})$. The rate that ψ changes for the material element is then given by the component of its velocity in the direction of the gradient of ψ times that gradient, namely by $\mathbf{v} \cdot \nabla \psi$.

10.4 REYNOLDS' TRANSPORT THEOREM

Owing to the kinematic relationship between Lagrangian and Eulerian descriptions, the material derivative emerges naturally in the laws governing field properties. Changes of ψ described by (10.10) apply to an infinitesimal material element. Consider now changes for a finite material volume $V(t)$ (i.e., one containing a fixed collection of matter). Because $V(t)$ has finite dimension, we must account for

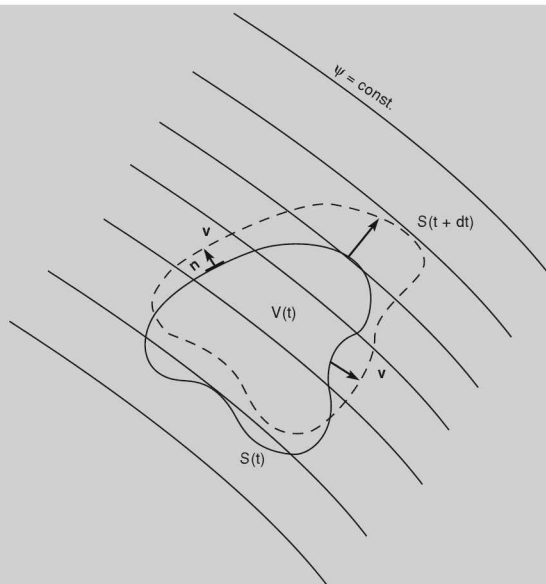


Figure 10.5 Finite material volume $V(t)$, containing a fixed collection of matter, that is displaced across a field property ψ .

variations of velocity \mathbf{v} and property ψ across the material volume. Consider the integral property

$$\int_{V(t)} \psi(x, y, z, t) dV'$$

over the finite system. The time rate of change of this property follows by differentiating the volume integral. However, the time derivative $\frac{d}{dt}$ cannot be commuted inside the integral because the limits of integration (viz. the position and form of the material volume) are themselves variable.

The time rate of change of the integral material property has two contributions (Fig. 10.5), analogous to those treated above for an infinitesimal material element:

- (1) Values of $\psi(\mathbf{x}, t)$ within the instantaneous material volume change temporally due to unsteadiness of the field.
- (2) The material volume moves to regions of different field values. Relative to a frame moving with $V(t)$, such motion introduces a flux of property ψ across the material volume's surface $S(t)$.

The first contribution is just the collective time rate of change of $\psi(\mathbf{x}, t)$ within the material volume:

$$\int_{V(t)} \frac{\partial \psi}{\partial t} dV'.$$

The second is the net transfer of ψ across the boundary of $V(t)$. If $S(t)$ has the local outward normal \mathbf{n} and local velocity \mathbf{v} , the local flux of ψ relative to that section of material surface is $-\psi \mathbf{v}$. Then the local flux of ψ into the material volume is given

by $-\psi \mathbf{v} \cdot \mathbf{n} = \psi \mathbf{v} \cdot \mathbf{n}$. Integrating over $S(t)$ gives the net rate that ψ is transferred into the material volume $V(t)$ due to its motion across contours of ψ :

$$\int_{S(t)} \psi \mathbf{v} \cdot \mathbf{n} dS'.$$

Collecting the two contributions and applying Gauss' theorem obtains

$$\begin{aligned} \frac{d}{dt} \int_{V(t)} \psi dV' &= \int_{V(t)} \left\{ \frac{\partial \psi}{\partial t} + \nabla \cdot (\mathbf{v}\psi) \right\} dV' \\ &= \int_{V(t)} \left\{ \frac{d\psi}{dt} + \psi \nabla \cdot \mathbf{v} \right\} dV', \end{aligned} \quad (10.11)$$

for the time rate of change of the integral material property. Known as *Reynolds' transport theorem*, (10.11) relates the time rate of change of some property of a finite body of fluid to the corresponding field variable and to the velocity field $\mathbf{v}(\mathbf{x}, t)$. As such, it constitutes a transformation between the Lagrangian and Eulerian descriptions of fluid motion.¹ To develop the equations that govern field variables in the Eulerian description of atmospheric motion, we apply Reynolds' theorem to properties of an arbitrary material volume.

10.5 CONSERVATION OF MASS

Let $\psi = \rho(\mathbf{x}, t)$. An arbitrary material volume $V(t)$ then has mass

$$\int_{V(t)} \rho(\mathbf{x}, t) dV'.$$

Because that system is comprised of a fixed collection of matter, the time rate of change of its mass must vanish

$$\frac{d}{dt} \int_{V(t)} \rho(\mathbf{x}, t) dV' = 0. \quad (10.12)$$

Applying Reynolds' transport theorem transforms (10.12) into

$$\int_{V(t)} \left\{ \frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} \right\} dV' = 0. \quad (10.13)$$

This relation must hold for "arbitrary" material volume $V(t)$. It follows that the integrand must vanish identically.

Conservation of mass for individual bodies of air thus requires

$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} = 0 \quad (10.14.1)$$

or, in flux form,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (10.14.2)$$

Known as the *continuity equation*, (10.14) provides a constraint on the field variables $\rho(\mathbf{x}, t)$ and $\mathbf{v}(\mathbf{x}, t)$. It must be satisfied throughout the domain and continuously in time.

¹ Mathematically, the transport theorem is a generalization to a moving body of fluid of Leibniz' rule for differentiating an integral with variable limits. It enables the time derivative of the integral property to be expressed in terms of an integral of time and space derivatives.

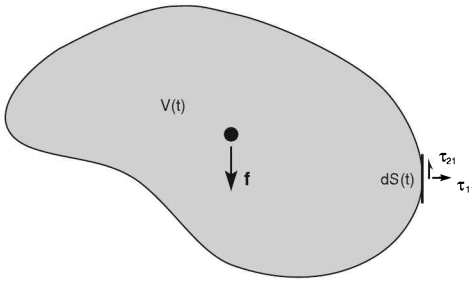


Figure 10.6 Finite material volume $V(t)$ that experiences a specific body force \mathbf{f} internally and a stress tensor $\boldsymbol{\tau}$ on its surface.

If material volume does not change for individual bodies of air, the motion is said to be *incompressible*. The specific volume $\alpha = 1/\rho$ for an arbitrary material element is then constant. Under these conditions, (10.14.1) reduces to

$$\nabla \cdot \mathbf{v} = 0,$$

which provides a simpler constraint on the motion.

Consider a specific field property f (viz. one referenced to a unit mass). Then $\psi = \rho f$ represents the absolute concentration of that property. Reynolds' transport theorem implies

$$\begin{aligned} \frac{d}{dt} \int_{V(t)} \rho f dV' &= \int_{V(t)} \left\{ \frac{d}{dt} (\rho f) + \rho f (\nabla \cdot \mathbf{v}) \right\} dV' \\ &= \int_{V(t)} \left\{ \rho \frac{df}{dt} + \left[\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} \right] \right\} dV'. \end{aligned} \quad (10.15)$$

By the continuity equation, the term in square brackets vanishes. Thus (10.15) reduces to the identity

$$\frac{d}{dt} \int_{V(t)} \rho f dV' = \int_{V(t)} \rho \frac{df}{dt} dV'. \quad (10.16)$$

For an absolute concentration, the time rate of change of an integral material property assumes this simpler form.

10.6 THE MOMENTUM BUDGET

10.6.1 Cauchy's Equations of Motion

In an inertial reference frame, Newton's second law of motion applied to the material volume $V(t)$ may be expressed

$$\frac{d}{dt} \int_{V(t)} \rho \mathbf{v} dV' = \int_{V(t)} \rho \mathbf{f} dV' + \int_{S(t)} \boldsymbol{\tau} \cdot \mathbf{n} dS', \quad (10.17)$$

where $\rho \mathbf{v}$ is the absolute concentration of momentum, \mathbf{f} is the specific body force acting internal to the material volume, and $\boldsymbol{\tau}$ is the *stress tensor* acting on its surface (Fig. 10.6). The stress tensor $\boldsymbol{\tau}$ is a counterpart of the deformation tensor \mathbf{e} . It represents the vector force per unit area exerted on surfaces normal to the three coordinate directions. Then $\boldsymbol{\tau} \cdot \mathbf{n}$ is the vector force per unit area exerted on the section of material surface with unit normal \mathbf{n} .

For common fluids such as air and water, the stress tensor is symmetric

$$\tau_{ji} = \tau_{ij}. \quad (10.18)$$

Like $\mathbf{e}(\mathbf{x}, t)$, the local stress tensor contains two basic contributions. Diagonal components of $\boldsymbol{\tau}(\mathbf{x}, t)$ define *normal stresses*. They act orthogonal to surfaces with normals in the three coordinate directions. The component $\tau_{11}(\mathbf{x}, t)$ describes the force per unit area acting in the direction of coordinate 1 on an element of surface with normal in the same direction (Fig. 10.6). Off-diagonal components of $\boldsymbol{\tau}(\mathbf{x}, t)$ define *shear stresses* that act tangential to those surfaces. The component τ_{21} describes the force per unit area acting in the direction of coordinate 2 on an element of surface with normal in the direction of coordinate 1.

Each of the stresses in $\boldsymbol{\tau}$ also represents a flux of momentum. The normal stress τ_{ii} (summation suspended) represents the flux of i momentum in the i direction. The shear stress τ_{ji} represents the flux of j momentum in the i direction. Formally, these fluxes are accomplished by molecular diffusion of momentum, which gives fluid viscosity. However, in large-scale atmospheric motion, similar fluxes of momentum are accomplished through mixing by small-scale turbulence. Much faster than molecular diffusion, turbulent fluxes are treated in analogous fashion. Both render the behavior of an individual material element diabatic because they transfer heat and momentum across its boundary (see Fig. 2.1).

In the atmosphere, the body force and stress tensor are either prescribed or are determined by the velocity field. Incorporating Reynolds' transport theorem for an absolute concentration (10.16) and Gauss' theorem transforms (10.17) into

$$\int_{V(t)} \rho \frac{d\mathbf{v}}{dt} dV' = \int_{V(t)} \{\rho \mathbf{f} + \nabla \cdot \boldsymbol{\tau}\} dV'$$

or

$$\int_{V(t)} \left\{ \rho \frac{d\mathbf{v}}{dt} - \rho \mathbf{f} - \nabla \cdot \boldsymbol{\tau} \right\} dV' = 0. \quad (10.19)$$

As before, (10.19) must hold for arbitrary material volume. The integrand must then vanish identically. Thus Newton's second law for individual bodies of air requires

$$\rho \frac{d\mathbf{v}}{dt} = \rho \mathbf{f} + \nabla \cdot \boldsymbol{\tau}. \quad (10.20)$$

Known as *Cauchy's equations*, (10.20) provide constraints on the field properties $\rho(\mathbf{x}, t)$ and $\mathbf{v}(\mathbf{x}, t)$. They must hold pointwise throughout the domain and continuously in time.²

The body force relevant to the atmosphere is gravity

$$\mathbf{f} = \mathbf{g}, \quad (10.21)$$

which is prescribed. On the other hand, internal stresses are determined autonomously by the motion. A *Newtonian fluid* like air has a stress tensor that is linearly proportional to the local rate of strain $\mathbf{e}(\mathbf{x}, t)$. In the absence of motion, $\boldsymbol{\tau}$ reduces to the normal stresses exerted by pressure. For our purposes, it suffices

² In the absence of friction, these are called *Euler's equations*.

to define the stress tensor as

$$\boldsymbol{\tau} = \begin{bmatrix} -p & & & \\ & 2\mu e_{ij} & & \\ & & -p & \\ & 2\mu e_{ji} & & \\ & & & -p \end{bmatrix}, \quad (10.22)$$

where μ is the *coefficient of viscosity*.³ Equation (10.22) is the form assumed by the stress tensor for an incompressible fluid. Shear stresses internal to the fluid then arise from shear in the motion field $\frac{\partial v_i}{\partial x_j}$. When expressed in terms of specific momentum, the governing equations involve the *kinematic viscosity* $\nu = \frac{\mu}{\rho}$. Also called the molecular *diffusivity*, ν has dimensions $\frac{\text{length}}{\text{time}^2}$. Turbulent momentum transfer is frequently modeled as diffusion. An eddy diffusivity is then used in place of ν (Chap. 13).

With the stress tensor $\boldsymbol{\tau}$ defined by (10.22), Cauchy's equations of motion reduce to

$$\frac{d\mathbf{v}}{dt} = \mathbf{g} - \frac{1}{\rho} \nabla p - \mathbf{D}, \quad (10.23.1)$$

where

$$\begin{aligned} \mathbf{D} &= -\frac{1}{\rho} \nabla \cdot \boldsymbol{\tau} \\ &= -\frac{1}{\rho} \nabla \cdot (\mu \nabla \mathbf{v}) = -\frac{1}{\rho} \frac{\partial}{\partial x_j} \mu \frac{\partial v_i}{\partial x_j} \end{aligned} \quad (10.23.2)$$

denotes the specific *drag force* exerted on the material element at location \mathbf{x} . Known as the *momentum equations*, (10.23) are a simplified form of the *Navier Stokes equations*. The latter embody the full representation of $\boldsymbol{\tau}$ in terms of \mathbf{e} and constitute the formal description of fluid motion. Conceptually, (10.23) asserts that the momentum of a material element changes according to the resultant force exerted on it by gravity, pressure gradient, and frictional drag.

10.6.2 Momentum equations in a rotating reference frame

Because they follow from Newton's laws of motion, the momentum equations apply in an inertial reference frame. However, the reference frame of the Earth (in which atmospheric motion is observed) is rotating and therefore noninertial. To apply in that reference frame, the momentum equations must be modified. Scalar quantities like $\rho(\mathbf{x}, t)$ appear the same in inertial and noninertial reference frames. So do their Lagrangian derivatives. However, vector quantities differ between those reference frames. Vector variables that describe a material element's motion (e.g., \mathbf{x} , $\mathbf{v} = \frac{d\mathbf{x}}{dt}$, and $\mathbf{a} = \frac{d\mathbf{v}}{dt}$) must therefore be corrected to account for acceleration of the Earth's reference frame.

Consider a reference frame rotating with angular velocity $\boldsymbol{\Omega}$ (Fig. 10.7). A vector \mathbf{A} which is constant in that frame rotates when viewed in an inertial reference frame.

³ Strictly, normal stresses also include a contribution from viscosity, but that effect is small enough to be neglected for most applications (e.g., Aris, 1962).

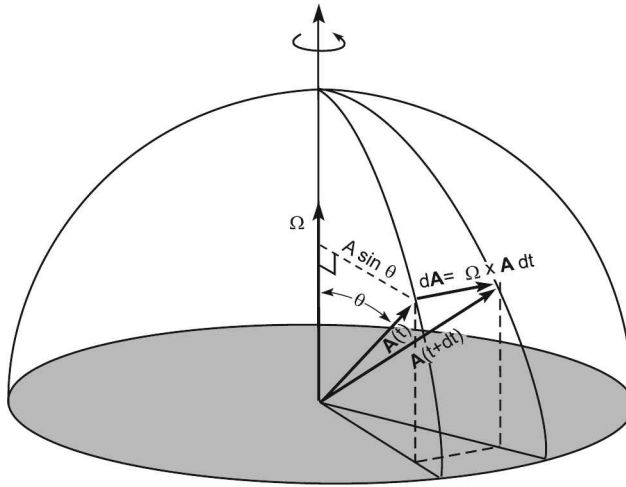


Figure 10.7 A vector \mathbf{A} that is fixed in a rotating reference frame changes in an inertial reference frame during an interval dt by the increment $|d\mathbf{A}| = A\sin\theta \cdot \Omega dt$ in a direction orthogonal to the plane of \mathbf{A} and $\boldsymbol{\Omega}$, or by the vector increment $d\mathbf{A} = \boldsymbol{\Omega} \times \mathbf{A} dt$.

During an interval dt , \mathbf{A} changes by a vector increment $d\mathbf{A}$ which is perpendicular to the plane of \mathbf{A} and $\boldsymbol{\Omega}$ and which has magnitude

$$|d\mathbf{A}| = A\sin\theta \cdot \Omega dt,$$

where θ is the angle between \mathbf{A} and $\boldsymbol{\Omega}$. Hence, from an inertial reference frame, the vector \mathbf{A} changes at the rate

$$\left| \frac{d\mathbf{A}}{dt} \right| = A\Omega\sin\theta$$

and in a direction perpendicular to the plane of \mathbf{A} and $\boldsymbol{\Omega}$. It follows that the time rate of change of \mathbf{A} apparent in an inertial reference frame is described by

$$\left(\frac{d\mathbf{A}}{dt} \right)_i = \boldsymbol{\Omega} \times \mathbf{A}. \quad (10.24)$$

More generally, a vector \mathbf{A} that has the time rate of change $\frac{d\mathbf{A}}{dt}$ in a rotating reference frame has the time rate of change

$$\left(\frac{d\mathbf{A}}{dt} \right)_i = \frac{d\mathbf{A}}{dt} + \boldsymbol{\Omega} \times \mathbf{A} \quad (10.25)$$

in an inertial reference frame.

Consider the position \mathbf{x} of a material element. By (10.25), the element's velocity $\mathbf{v} = \frac{d\mathbf{x}}{dt}$ apparent in an inertial reference frame is

$$\mathbf{v}_i = \mathbf{v} + \boldsymbol{\Omega} \times \mathbf{x}. \quad (10.26)$$

Similarly, the acceleration that is apparent in the inertial frame is given by

$$\left(\frac{d\mathbf{v}_i}{dt} \right)_i = \frac{d\mathbf{v}_i}{dt} + \boldsymbol{\Omega} \times \mathbf{v}_i.$$

Incorporating the velocity apparent in the inertial frame (10.26) obtains

$$\begin{aligned} \left(\frac{d\mathbf{v}_i}{dt}\right)_i &= \left(\frac{d\mathbf{v}}{dt} + \boldsymbol{\Omega} \times \mathbf{v}\right) + \boldsymbol{\Omega} \times (\mathbf{v} + \boldsymbol{\Omega} \times \mathbf{x}) \\ &= \frac{d\mathbf{v}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{v} + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{x}) \end{aligned} \quad (10.27)$$

for the acceleration apparent in the inertial frame.

According to (10.27), rotation of the reference frame introduces two corrections to the material acceleration. The first, $2\boldsymbol{\Omega} \times \mathbf{v}$, is the *Coriolis acceleration*. Perpendicular to an air parcel's velocity and to the *planetary vorticity* $2\boldsymbol{\Omega}$, it represents an apparent force in the rotating frame of the Earth. The Coriolis acceleration is important for motions that have time scales comparable to that of the Earth's rotation (Chap. 12). The second correction, $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{x})$, is the *centrifugal acceleration* of an air parcel. Likewise an apparent force, it was treated in Chap. 6. When geopotential coordinates are used, this apparent force is absorbed into effective gravity.

Incorporating the material acceleration apparent in an inertial reference frame (10.27) transforms the momentum equations (10.23) into a form valid in the rotating frame of the Earth:

$$\frac{d\mathbf{v}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{1}{\rho} \nabla p - g\mathbf{k} - \mathbf{D}, \quad (10.28)$$

where g is understood to denote effective gravity and \mathbf{k} the upward normal in the direction of increasing geopotential (i.e., $g\mathbf{k} = \nabla\Phi$). The correction $2\boldsymbol{\Omega} \times \mathbf{v}$ represents the "Coriolis acceleration" when it appears on the left-hand side of the momentum balance. If moved to the right-hand side, it represents the "Coriolis force" $-2\boldsymbol{\Omega} \times \mathbf{v}$, an apparent force that acts on an air parcel in the rotating frame of the Earth. Because it acts orthogonal to the parcel's displacement, the Coriolis force performs no work.

10.7 THE FIRST LAW OF THERMODYNAMICS

Applied to the material volume $V(t)$, the First Law may be expressed

$$\frac{d}{dt} \int_{V(t)} \rho c_v T dV' = - \int_{S(t)} \mathbf{q} \cdot \mathbf{n} dS' - \int_{V(t)} \rho p \frac{d\alpha}{dt} dV' + \int_{V(t)} \rho \dot{q} dV', \quad (10.29)$$

where $c_v T$ represents the specific internal energy. Forcing it on the right-hand side are three terms: (1) \mathbf{q} is the local heat flux, so $-\mathbf{q} \cdot \mathbf{n}$ represents the heat flux "into" the material volume. (2) $\alpha = \frac{1}{\rho}$ is the specific volume, so $p \frac{d\alpha}{dt}$ represents the specific work rate. (3) \dot{q} denotes the specific rate of internal heating (e.g., associated with the latent heat release and frictional dissipation of motion).

The local rate of expansion work is related to the dilatation of the material element occupying that position. Because

$$\begin{aligned} \frac{1}{\rho} \frac{d\rho}{dt} &= \alpha \frac{d\left(\frac{1}{\alpha}\right)}{dt} \\ &= -\frac{1}{\alpha} \frac{d\alpha}{dt}, \end{aligned} \quad (10.30)$$

the continuity equation (10.14) may be expressed

$$\frac{1}{\alpha} \frac{d\alpha}{dt} = \nabla \cdot \mathbf{v}. \quad (10.31)$$

Incorporating (10.31), along with Reynolds' transport theorem (10.16) and Gauss' theorem, transforms (10.29) into

$$\int_{V(t)} \left\{ \rho c_v \frac{dT}{dt} + \nabla \cdot \mathbf{q} + p \nabla \cdot \mathbf{v} - \rho \dot{q} \right\} dV' = 0.$$

Again, because $V(t)$ is arbitrary, the integrand must vanish identically.

Therefore, the First Law applied to individual bodies of air requires

$$\rho c_v \frac{dT}{dt} = -\nabla \cdot \mathbf{q} - p \nabla \cdot \mathbf{v} + \rho \dot{q}, \quad (10.32)$$

which provides a constraint on the field variables involved. Of the three thermodynamic properties represented, only two are independent because air behaves as a pure substance (Sec. 2.1.4); they must also satisfy the gas law. Similarly, the heat flux \mathbf{q} and the internal heating rate \dot{q} are determined autonomously by properties already represented. Therefore, they introduce no additional unknowns.

It is convenient to separate the heat flux into radiative and diffusive components:

$$\begin{aligned} \mathbf{q} &= \mathbf{q}_R + \mathbf{q}_T \\ &= \mathbf{F} - k \nabla T, \end{aligned} \quad (10.33)$$

where \mathbf{F} is the net radiative flux (Sec. 8.2) and k denotes the thermal conductivity in Fourier's law of heat conduction. The first law then becomes

$$\rho c_v \frac{dT}{dt} + p \nabla \cdot \mathbf{v} = -\nabla \cdot \mathbf{F} + \nabla \cdot (k \nabla T) + \rho \dot{q}. \quad (10.34)$$

Known as the *thermodynamic equation*, (10.34) expresses the rate that a material element's internal energy changes in terms of the rate that work is performed on it and the net rate it absorbs heat through convergence of radiative and diffusive energy fluxes.

The thermodynamic equation can be expressed more compactly in terms of potential temperature. For an individual air parcel, the fundamental relation

$$du = T ds - p d\alpha \quad (10.35.1)$$

relates the change of entropy to other material properties. Incorporating the identity between entropy and potential temperature (3.25) transforms this into

$$c_p T d \ln \theta = du + p d\alpha. \quad (10.35.2)$$

Differentiating with respect to time, with the material coordinate ξ fixed, obtains

$$c_p T \frac{d \ln \theta}{dt} = \frac{du}{dt} + p \frac{d\alpha}{dt}, \quad (10.36)$$

where $\frac{d}{dt}$ represents the Lagrangian derivative (10.10). Multiplying by ρ and introducing the continuity equation (10.4) transforms this into

$$\frac{\rho c_p T}{\theta} \frac{d\theta}{dt} = \rho c_v \frac{dT}{dt} + p \nabla \cdot \mathbf{v}. \quad (10.37)$$

Then incorporating (10.37) into (10.34) absorbs the compression work into the time rate of change of θ , yielding the thermodynamic equation

$$\rho \frac{c_p T}{\theta} \frac{d\theta}{dt} = -\nabla \cdot \mathbf{F} + \nabla \cdot (k \nabla T) + \rho \dot{q}. \quad (10.38)$$

Collectively, the continuity, momentum, and thermodynamic equations represent five partial differential equations in five independent field variables: three components of motion and two independent thermodynamic properties. Advective contributions to the material derivative, like advection of momentum $\mathbf{v} \cdot \nabla \mathbf{v}$ and of temperature $\mathbf{v} \cdot \nabla T$, make the governing equations nonlinear – quadratically. Their solution requires initial conditions that specify the preliminary state of the atmosphere and boundary conditions that specify its properties along physical borders. Referred to as *the equations of motion*, these equations govern the behavior of a compressible, stratified atmosphere in a rotating reference frame. As such, they constitute the starting point for dynamical investigations, as well as for investigations of chemistry and radiation in the presence of motion.

SUGGESTED REFERENCES

Vectors, Tensors, and the Basic Equations of Fluid Mechanics (1962) by Aris develops the kinematics of fluid motion, along with the governing equations from a Lagrangian perspective.

An Introduction to Dynamic Meteorology (2004) by Holton contains alternate derivations of the continuity and thermodynamic equations.

PROBLEMS

1. Consider the 2-dimensional motion

$$u(x, y, t) = \frac{\partial \psi}{\partial y}$$

$$v(x, y, t) = -\frac{\partial \psi}{\partial x},$$

where

$$\psi(x, y, t) = 4e^{-\left\{(x-x_0(t))^2 + (y-y_0(t))^2\right\}}$$

defines a family of streamlines and

$$x_0(t) = y_0(t) = t.$$

Plot the parcel trajectory and streakline from $t = 0$ to 4 beginning at (a) $(x, y) = (0, 1)$, (b) $(x, y) = (-1, 0)$, (c) $(x, y) = (0, 0)$.

2. For the motion in Prob. 10.1, plot at $t = 1, 2, 3, 4$ the material volume which, at $t = 0$, is defined by the radial coordinates:

$$1 - \Delta r < r < 1 + \Delta r$$

$$-\Delta \phi < \phi < \Delta \phi$$

for (a) $\Delta r = 0.5$ and $\Delta \phi = \frac{\pi}{4}$, (b) $\Delta r = 0.1$ and $\Delta \phi = 0.1$. (c) Contrast the deformations experienced by the material volumes in (a) and (b) and use them to infer the limiting behavior for Δr and $\Delta \phi \rightarrow 0$.

3. Simplify the equations of motion for the special case of (a) incompressible motion, wherein the volume of a material element is conserved, (b) adiabatic motion.
4. Describe the circumstances under which a property ψ is conserved, yet varies spatially in a steady flow.
5. A free surface is one that moves to alleviate any stress and maintain $\tau = 0$. If the rate of strain tensor is dominated by the vertical shears $\frac{\partial u}{\partial z}$ and $\frac{\partial v}{\partial z}$, describe how the horizontal velocity varies adjacent to a free surface.
6. A geostationary satellite is positioned over latitude $\phi_0 = 30^\circ$ and longitude $\lambda_0 = 0^\circ$. From it, a projectile is fired northward at a speed v_0 . By ignoring the satellite's altitude, calculate the longitude where the projectile crosses 45° latitude if (a) $v_0 = 1000 \text{ m s}^{-1}$, (b) $v_0 = 100 \text{ m s}^{-1}$, (c) $v_0 = 10 \text{ m s}^{-1}$. (d) For each case, evaluate the dimensionless time for the traversal, scaled by the rotation period of the Earth.

7. Show that the antisymmetric component of the velocity gradient tensor may be expressed in terms of the angular velocity vector $\boldsymbol{\omega}$ in (10.7).
8. Show that the vorticity equals twice the local angular velocity (10.8).
9. Use the vector identity (D.14) in Appendix D to show that the material acceleration may be expressed

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + \nabla \frac{|\mathbf{v}|^2}{2} + \boldsymbol{\zeta} \times \mathbf{v},$$

where $\boldsymbol{\zeta} = \nabla \times \mathbf{v}$.

10. Demonstrate that the Coriolis force does not enter the budget of specific kinetic energy $\frac{|\mathbf{v}|^2}{2}$.
11. Show that, for 2-dimensional nondivergent motion in an inertial reference frame, the vorticity $\zeta = \mathbf{k} \cdot \nabla \times \mathbf{v}$ is a conserved property.
12. Consider a parcel with local speed v , the natural coordinate s measured along the parcel's trajectory, and a unit vector \mathbf{s} that is everywhere tangential to the trajectory. (a) Express the material derivative in terms of v , s , and \mathbf{s} . (b) Show that the material acceleration described by

$$\frac{d\mathbf{v}}{dt} = \frac{dv}{dt} \mathbf{s} + v \frac{d\mathbf{s}}{dt}.$$

(c) Interpret the two accelerations appearing on the right.